# HEALTHCARE ANALYSIS

By: Jermaine Sangiwa

# BACKGROUND INFORMATIOIN

- As a Healthcare Analyst, I've been asked to analyse the hospital records to track the progress of some common illnesses so that the public health programme is well informed.

# DATASET

Patients

| patient_id | patient_name | age | gender | city |
|---|---|---|---|---|
| 1 | John Smith | 45 | Male | Seattle |
| 2 | Jane Doe | 32 | Female | Miami |
| 3 | Mike Johnson | 50 | Male | Seattle |
| 4 | Lisa Jones | 28 | Female | Miami |
| 5 | David Kim | 60 | Male | Chicago |

Symptoms

| symptom_id | symptom_name |
|---|---|
| 1 | Fever |
| 2 | Cough |
| 3 | Difficulty Breathing |
| 4 | Fatigue |
| 5 | Headache |

Diagnoses

| diagnosis_id | diagnosis_name |
|---|---|
| 1 | Common Cold |
| 2 | Influenza |
| 3 | Pneumonia |
| 4 | Bronchitis |
| 5 | COVID-19 |

Visits

| visit_id | patient_id | symptom_id | diagnosis_id | visit_date |
|---|---|---|---|---|
| 1 | 1 | 1 | 2 | 1/1/22 |
| 2 | 2 | 2 | 1 | 2/1/22 |
| 3 | 3 | 3 | 3 | 2/1/22 |
| 4 | 4 | 1 | 4 | 3/1/22 |
| 5 | 5 | 2 | 5 | 3/1/22 |
| 6 | 1 | 4 | 1 | 13/5/22 |
| 7 | 3 | 4 | 1 | 20/5/22 |
| 8 | 3 | 2 | 1 | 20/5/22 |
| 9 | 2 | 1 | 4 | 19/8/22 |
| 10 | 1 | 2 | 5 | 1/12/22 |

```sql
SELECT
    Patients.patient_id,
    Patients.patient_name AS Patient_Name

FROM Visits
JOIN Patients
JOIN Diagnoses
ON Visits.patient_id = Patients.patient_id
AND Visits.diagnosis_id = Diagnoses.diagnosis_id

WHERE
    Diagnoses.diagnosis_name = 'COVID-19'

ORDER BY
    Patients.patient_id
```

| | patient_id | Patient_Name |
|---|---|---|
| 1 | 1 | John Smith |
| 2 | 5 | David Kim |

# QUESTION 1: WRITE A SQL QUERY TO RETRIEVE ALL PATIENTS WHO HAVE BEEN DIAGNOSED WITH ‘COVID-19’.

```sql
SELECT
    Patients.patient_id,
    Patients.patient_name AS Patient_Name,
    COUNT(Visits.visit_id) AS Number_of_Visits

FROM Visits
JOIN Patients
ON Visits.patient_id = Patients.patient_id

GROUP BY Patients.patient_id
ORDER BY Number_of_Visits DESC
```

| patient_id | Patient_Name | Number_of_Visits |
|---|---|---|
| 3 | Mike Johnson | 3 |
| 1 | John Smith | 3 |
| 2 | Jane Doe | 2 |
| 5 | David Kim | 1 |
| 4 | Lisa Jones | 1 |

# QUESTION 2: WRITE A SQL QUERY TO RETRIEVE THE NUMBER OF VISITS MADE BY EACH PATIENTS, ORDERED BY THE NUMBER OF VISITS IN DESCENDING ORDER

```sql
SELECT
    AVG(Patients.age) AS Average_Age

FROM Visits
JOIN Patients
JOIN Diagnoses
ON Visits.patient_id = Patients.patient_id
AND Visits.diagnosis_id = Diagnoses.diagnosis_id

WHERE
    Diagnoses.diagnosis_name = 'Pneumonia'
```

| Average_Age |
| --- |
| 50.0 |

# QUESTION 3: WRITE A SQL QUERY TO CALCULATE THE AVERAGE AGE OF PATIENT WHO HAVE BEEN DIAGNOSED WITH PNEUMONIA.

```sql
SELECT
    Symptoms.symptom_id,
    Symptoms.symptom_name,
    COUNT(Visits.visit_id) AS Number_of_Visits

FROM Visits
JOIN Symptoms
ON Visits.symptom_id = Symptoms.symptom_id

GROUP BY Symptoms.symptom_id
ORDER BY Number_of_Visits DESC
LIMIT 3
```

| symptom_id | symptom_name | Number_of_Visits |
|---|---|---|
| 2 | Cough | 4 |
| 1 | Fever | 3 |
| 4 | Fatigue | 2 |

# QUESTION 4 WRITE A SQL QUERY TO RETRIEVE THE TOP 3 MOST COMMON SYMPTOMS AMONG ALL VISITS.

```sql
SELECT
    Patients.patient_id,
    Patients.patient_name,
    COUNT(DISTINCT Visits.symptom_id) AS Number_of_Symptoms

FROM Visits
JOIN Patients
ON Visits.patient_id = Patients.patient_id

GROUP BY Patients.patient_id
ORDER BY Number_of_Symptoms DESC
```

| patient_id | patient_name | Number_of_Symptoms |
|---|---|---|
| 3 | Mike Johnson | 3 |
| 1 | John Smith | 3 |
| 2 | Jane Doe | 2 |
| 5 | David Kim | 1 |
| 4 | Lisa Jones | 1 |

# QUESTION 5: WRITE A SQL QUERY TO RETRIEVE THE PATIENT WHO HAS THE HIGHEST NUMBER OF DIFFERENT SYMPTOMS REPORTED.

```sql
SELECT
    SUM(CASE WHEN Diagnoses.diagnosis_name = 'COVID-19' THEN 1 ELSE 0 END) AS COVID_Patients,
    COUNT(DISTINCT Patients.patient_id) AS Total_Patients,
    (CAST(SUM(CASE WHEN Diagnoses.diagnosis_name = 'COVID-19' THEN 1 ELSE 0 END) AS FLOAT)/COUNT(DISTINCT Patients.patient_id))*100 AS Percentage

FROM Visits
JOIN Patients
JOIN Diagnoses
ON Visits.patient_id = Patients.patient_id
AND Visits.diagnosis_id = Diagnoses.diagnosis_id
```

| COVID_Patients | Total_Patients | Percentage |
|---|---|---|
| 2 | 5 | 40.0 |

# QUESTION 6: WRITE A SQL QUERY TO CALCULATE THE PERCENTAGE OF PATIENTS WHO HAVE BEEN DIAGNOSED WITH ‘COVID-19’ OUT OF THE TOTAL NUMBER OF PATIENTS.

```sql
SELECT
    Patients.city AS City,
    COUNT(Visits.visit_id) AS Number_of_Visits

FROM Visits
JOIN Patients
ON Visits.patient_id = Patients.patient_id

GROUP BY Patients.city
ORDER BY Number_of_Visits DESC
LIMIT 5
```

| City | Number_of_Visits |
|---|---|
| Seattle | 6 |
| Miami | 3 |
| Chicago | 1 |

# QUESTION 7: WRITE A SQL QUERY TO RETRIEVE THE TOP 5 CITIES WITH THE HIGHEST NUMBER OF VISITS, ALONG WITH THE COUNT OF VISITS IN EACH CITY.

```sql
SELECT
    Patients.patient_id,
    Patients.patient_name,
    Visits.visit_date,
    COUNT(Visits.visit_id) AS Visits_Number

FROM Visits
JOIN Patients
ON Visits.patient_id = Patients.patient_id

GROUP BY Patients.patient_id, Patients.patient_name, Visits.visit_date
ORDER BY Visits_Number DESC
```

| patient_id | patient_name | visit_date | Visits_Number |
|---|---|---|---|
| 3 | Mike Johnson | 2022-05-20 | 2 |
| 1 | John Smith | 2022-01-01 | 1 |
| 1 | John Smith | 2022-05-13 | 1 |
| 1 | John Smith | 2022-12-01 | 1 |
| 2 | Jane Doe | 2022-01-02 | 1 |
| 2 | Jane Doe | 2022-08-19 | 1 |
| 3 | Mike Johnson | 2022-01-02 | 1 |
| 4 | Lisa Jones | 2022-01-03 | 1 |
| 5 | David Kim | 2022-01-03 | 1 |

# QUESTION 8: WRITE A SQL QUERY TO FIND THE PATIENTS WITH THE HIGHEST NUMBER OF VISITS IN A SINGLE DAY, ALONG WITH THE CORRESPONDING VISIT DATE.

```sql
SELECT
    Diagnoses.diagnosis_id,
    Diagnoses.diagnosis_name,
    ROUND(AVG(Patients.age), 0) AS Average_Age

FROM Visits
JOIN Patients
JOIN Diagnoses
ON Visits.patient_id = Patients.patient_id
AND Visits.diagnosis_id - Diagnoses.diagnosis_id

GROUP BY Diagnoses.diagnosis_id
ORDER BY Average_Age DESC
```

| diagnosis_id | diagnosis_name | Average_Age |
|---|---|---|
| 4 | Bronchitis | 47.0 |
| 2 | Influenza | 44.0 |
| 1 | Common Cold | 43.0 |
| 3 | Pneumonia | 43.0 |
| 5 | COVID-19 | 42.0 |

# QUESTION 9: WRITE A SL QUERY TO RETRIEVE THE AVERAGE AGE OF PATIENTS FOR EACH DIAGNOSIS, ORDERED BY THE AVERAGE AGE IN DESCENDING ORDER.

```sql
SELECT
    v1.visit_date,
    COUNT(v2.visit_id) AS visit_number

FROM Visits AS v1
JOIN Visits AS v2
ON v1.visit_id = v2.visit_id

GROUP BY v1.visit_date
ORDER BY visit_number
```

| visit_date | visit_number |
|---|---|
| 2022-01-01 | 1 |
| 2022-05-13 | 1 |
| 2022-08-19 | 1 |
| 2022-12-01 | 1 |
| 2022-01-02 | 2 |
| 2022-01-03 | 2 |
| 2022-05-20 | 2 |

# QUESTION 10: WRITE A SQL QUERY TO CALCULATE THE CUMULATIVE COUNT OF VISITS OVER TIME, ORDERED BY THE VISIT DATE.

# THANK YOU