



SQL CASE STUDY

Healthcare Analysis (Challenge-7)

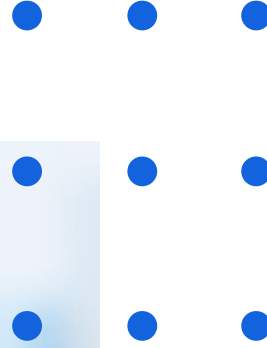
Presented by- Binoy Patra

Healthcare Analysis (Challenge-7)

Contents

- 1 Introduction
- 2 Tables
- 3 Questions and Answers





Introduction

As a Healthcare Analyst,I have been asked to analyse the hospital records to track the progress of some common illnesses so that our hospital can help to inform a public health programme.



Tables

Healthcare Analysis
(Challenge-7)

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

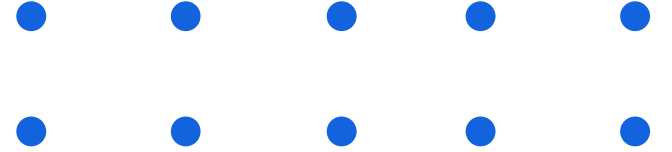
Visits

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





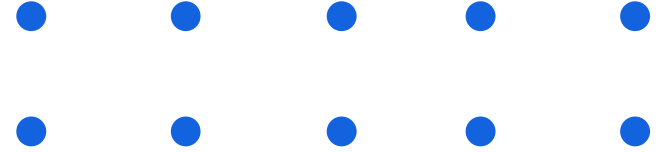
Healthcare Analysis (Challenge-7)



Questions & Answer



Healthcare Analysis (Challenge-7)



1. Write a SQL query to retrieve all patients who have been diagnosed with COVID-19.

```
SELECT
    P.patient_id, P.patient_name, D.diagnosis_name
FROM
    datacoach.patients P
    JOIN
    datacoach.visits V ON P.patient_id = V.patient_id
    JOIN
    datacoach.diagnoses D ON V.diagnosis_id = D.diagnosis_id
WHERE
    D.diagnosis_name = 'COVID-19'
ORDER BY P.patient_id;
```

	patient_id	patient_name	diagnosis_name
▶	1	John Smith	COVID-19
	5	David Kim	COVID-19



Healthcare Analysis (Challenge-7)



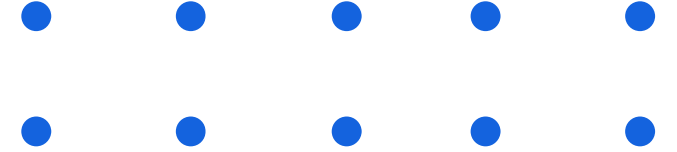
2. Write a SQL query to retrieve the number of visits made by each patient, ordered by the number of visits in descending order.

```
SELECT
    P.patient_id,
    P.patient_name,
    COUNT(V.patient_id) AS TimesVisited
FROM
    datacoach.patients P
    JOIN
    datacoach.visits V ON P.patient_id = V.patient_id
GROUP BY P.patient_id , P.patient_name
ORDER BY TimesVisited DESC;
```

	patient_id	patient_name	TimesVisited
▶	1	John Smith	3
	3	Mike Johnson	3
	2	Jane Doe	2
	4	Lisa Jones	1
	5	David Kim	1



Healthcare Analysis (Challenge-7)



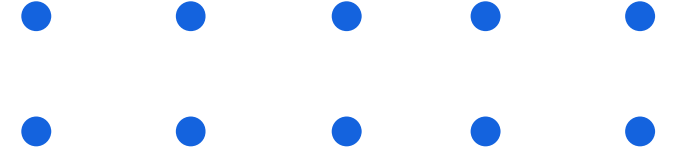
3. Write a SQL query to calculate the average age of patients who have been diagnosed with Pneumonia.

```
SELECT
    D.diagnosis_name, ROUND(AVG(P.age)) AS AverageAGE
FROM
    datacoach.patients P
    JOIN
    datacoach.visits V ON P.patient_id = V.patient_id
    JOIN
    datacoach.diagnoses D ON V.diagnosis_id = D.diagnosis_id
WHERE
    D.diagnosis_name = 'Pneumonia';
```

	diagnosis_name	AverageAGE
▶	Pneumonia	50



Healthcare Analysis (Challenge-7)



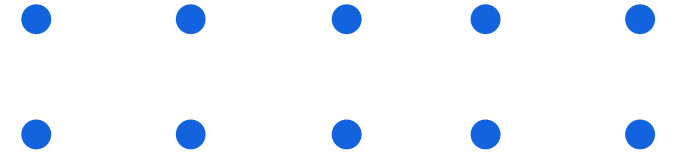
4. Write a SQL query to retrieve the top 3 most common symptoms among all visits.

```
SELECT
    S.symptom_name, COUNT(V.symptom_id) AS NumberOccured
FROM
    datacoach.symptoms S
    JOIN
    datacoach.visits V ON S.symptom_id = V.symptom_id
GROUP BY S.symptom_name
ORDER BY NumberOccured DESC
LIMIT 3;
```

	symptom_name	NumberOccured
►	Cough	4
	Fever	3
	Fatigue	2



Healthcare Analysis (Challenge-7)



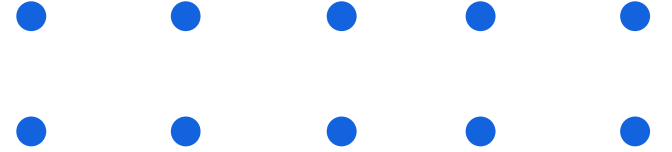
5. Write a SQL query to retrieve the patient who has the highest number of different symptoms reported.

```
WITH CTE1 AS (SELECT P.patient_id,P.patient_name ,COUNT(DISTINCT V.symptom_id) as Number_of_Symptoms,  
DENSE_RANK() OVER(ORDER BY COUNT(DISTINCT V.symptom_id) DESC) RN  
FROM datacoach.patients P  
JOIN datacoach.visits V ON P.patient_id=V.patient_id  
GROUP BY P.patient_id, P.patient_name)  
SELECT CTE1.PATIENT_ID,CTE1.PATIENT_NAME,CTE1.Number_of_Symptoms  
FROM CTE1  
WHERE CTE1.RN=1;
```

	PATIENT_ID	PATIENT_NAME	Number_of_Symptoms
▶	1	John Smith	3
	3	Mike Johnson	3



Healthcare Analysis (Challenge-7)



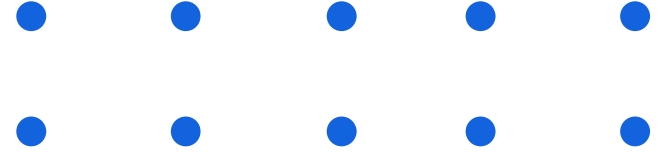
6. Write a SQL query to calculate the percentage of patients who have been diagnosed with COVID-19 out of the total number of patients.

```
WITH CTE1 AS (SELECT  
COUNT( CASE WHEN D.diagnosis_name= 'COVID-19' THEN V.patient_id END) AS Covid_Count,  
COUNT(DISTINCT(V.patient_id)) AS Total  
FROM datacoach.diagnoses D  
join datacoach.visits V ON D.diagnosis_id=V.diagnosis_id)  
SELECT CTE1.Covid_Count AS Covid_Patient, CTE1.Total AS Total_Patient,  
ROUND((CTE1.Covid_Count/CTE1.Total *100),2) AS 'Covid_%'  
from CTE1;
```

	Covid_Patient	Total_Patient	Covid_%
▶	2	5	40.00



Healthcare Analysis (Challenge-7)



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.

```
SELECT
    P.city, COUNT(V.patient_id) AS NumberVisited
FROM
    datacoach.visits V
    JOIN
    datacoach.patients P ON V.patient_id = P.patient_id
GROUP BY P.city;
```

	city	NumberVisited
▶	Seattle	6
	Miami	3
	Chicago	1



Healthcare Analysis (Challenge-7)



8. Write a SQL query to find the patient who has the highest number of visits in a single day, along with the corresponding visit date.

```
SELECT
    P.patient_id,
    P.patient_name,
    V.visit_date,
    COUNT(V.patient_id) NumberVisited
FROM
    datacoach.visits V
    JOIN
        datacoach.patients P ON V.patient_id = P.patient_id
GROUP BY P.patient_id , P.patient_name , V.visit_date
ORDER BY NumberVisited DESC
LIMIT 1;
```

	patient_id	patient_name	visit_date	NumberVisited
▶	3	Mike Johnson	2022-05-20	2



Healthcare Analysis (Challenge-7)



9. Write a SQL query to retrieve the average age of patients for each diagnosis, ordered by the average age in descending order.

```
SELECT
    D.diagnosis_name, ROUND(AVG(P.age)) AS AverageAge
FROM
    datacoach.patients P
    JOIN
    datacoach.visits V ON V.patient_id = P.patient_id
    JOIN
    datacoach.diagnoses D ON V.diagnosis_id = D.diagnosis_id
GROUP BY D.diagnosis_name
ORDER BY AverageAge DESC;
```

	diagnosis_name	AverageAge
►	COVID-19	53
	Pneumonia	50
	Influenza	45
	Common Cold	44
	Bronchitis	30



Healthcare Analysis (Challenge-7)



10. Write a SQL query to calculate the cumulative count of visits over time, ordered by the visit date.

```
SELECT V.visit_date, COUNT(V.visit_id) AS VISITOR,  
SUM(COUNT(V.patient_id)) OVER(ORDER BY V.visit_date) AS CumSumVisitor  
FROM datacoach.visits V  
GROUP BY V.visit_date ORDER BY V.visit_date;
```

	visit_date	VISITOR	CumSumVisitor
▶	2022-01-01	1	1
	2022-01-02	2	3
	2022-01-03	2	5
	2022-05-13	1	6
	2022-05-20	2	8
	2022-08-19	1	9
	2022-12-01	1	10

Healthcare Analysis (Challenge-7)

Thanks

❖ DataCoach

❖ Matthew Steel

