

Healthcare Data Insights: A Comprehensive SQL-Based Analysis

Project Objective

This project aims to analyze **healthcare data** using **SQL** to derive meaningful insights that can help hospitals, insurance providers, and medical professionals make **data-driven decisions**. The project covers essential areas such as **patient demographics, admission types, financial trends, hospital performance, and readmission risks**.

Link to Dataset: [Data](#)

Scope of the Analysis

This project is structured into three main sections:

1. **Basic Data Exploration:** Understanding patient distribution, hospital admissions, and common conditions.
2. **Financial and Insurance Analysis:** Assessing revenue, billing amounts, and insurance provider trends.
3. **Advanced Analysis & Predictive Insights:** Investigating patient readmissions, doctor performance, and multi-hospital visits.

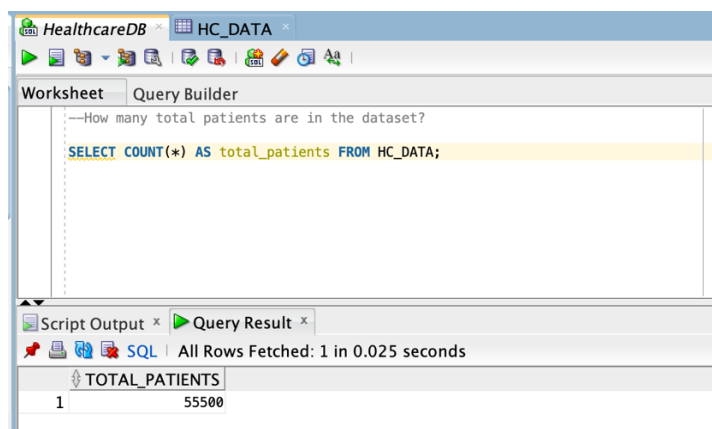
Data Insights & Query Explanations

Each query in this project serves a specific purpose and provides key insights:

1. Patient Demographics & Hospital Admissions

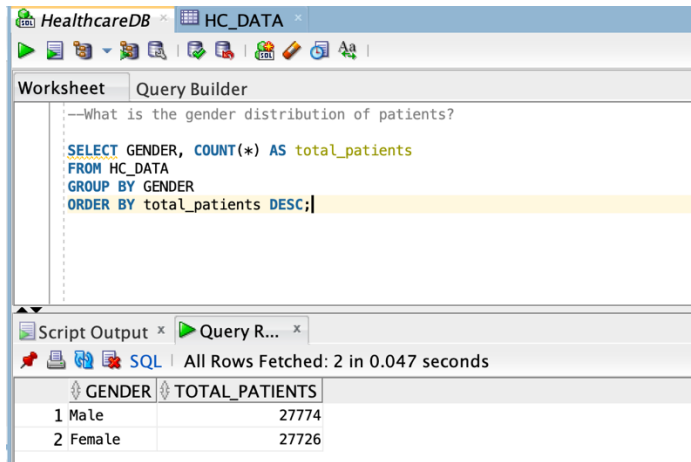
How many total patients are in the dataset?

- **Query:** We use `COUNT (*)` to get the total number of patient records.
- **Insight:** Helps in understanding the overall size of the dataset and hospital capacity planning.



What is the gender distribution of patients?

- **Query:** We use `GROUP BY GENDER` and `COUNT(*)` to count the number of patients per gender.
- **Insight:** Hospitals can analyze gender-based healthcare needs and ensure equitable treatment facilities.

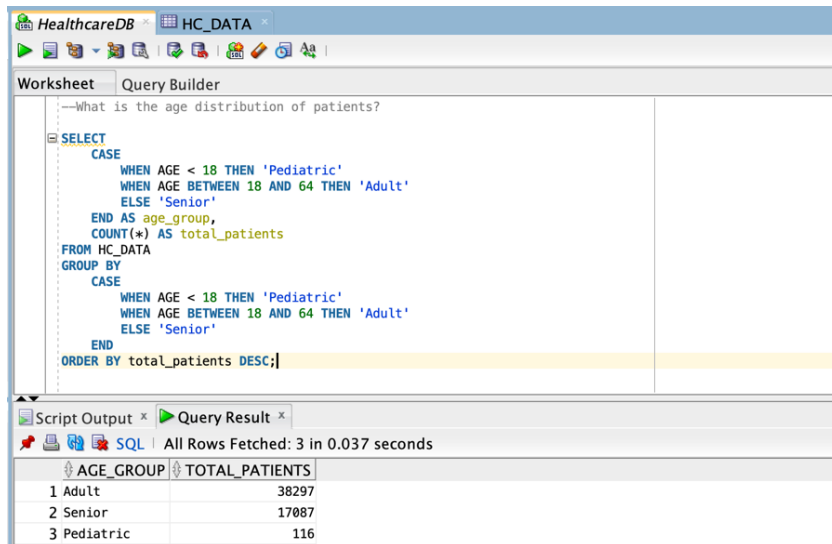


The screenshot shows a database query tool interface. The 'Query Builder' tab is active, displaying a query: `SELECT GENDER, COUNT(*) AS total_patients FROM HC_DATA GROUP BY GENDER ORDER BY total_patients DESC;`. Below the query, the 'Script Output' tab shows the results: 2 rows fetched in 0.047 seconds. The results are displayed in a table with columns 'GENDER' and 'TOTAL_PATIENTS'.

GENDER	TOTAL_PATIENTS
1 Male	27774
2 Female	27726

What is the age distribution of patients?

- **Query:** We classify patients into `Pediatric`, `Adult`, and `Senior` using a `CASE` statement and then count occurrences.
- **Insight:** Useful for planning age-specific healthcare programs and resource allocation.

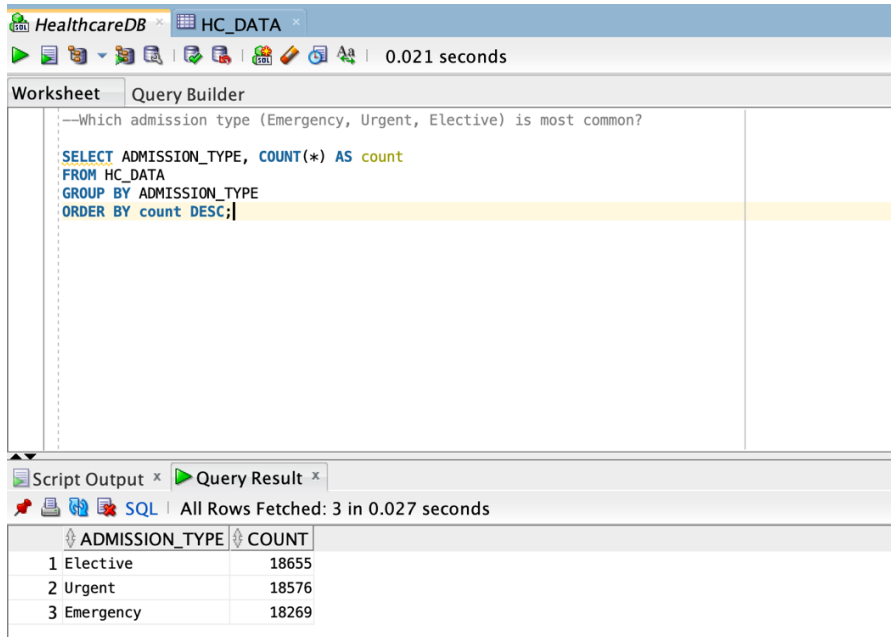


The screenshot shows a database query tool interface. The 'Query Builder' tab is active, displaying a query: `SELECT CASE WHEN AGE < 18 THEN 'Pediatric' WHEN AGE BETWEEN 18 AND 64 THEN 'Adult' ELSE 'Senior' END AS age_group, COUNT(*) AS total_patients FROM HC_DATA GROUP BY CASE WHEN AGE < 18 THEN 'Pediatric' WHEN AGE BETWEEN 18 AND 64 THEN 'Adult' ELSE 'Senior' END ORDER BY total_patients DESC;`. Below the query, the 'Script Output' tab shows the results: 3 rows fetched in 0.037 seconds. The results are displayed in a table with columns 'AGE_GROUP' and 'TOTAL_PATIENTS'.

AGE_GROUP	TOTAL_PATIENTS
1 Adult	38297
2 Senior	17087
3 Pediatric	116

Which admission type (Emergency, Urgent, Elective) is most common?

- **Query:** Using `GROUP BY ADMISSION_TYPE`, we count occurrences to find the most frequent type.
- **Insight:** Helps hospitals in optimizing emergency response systems and elective surgery scheduling.

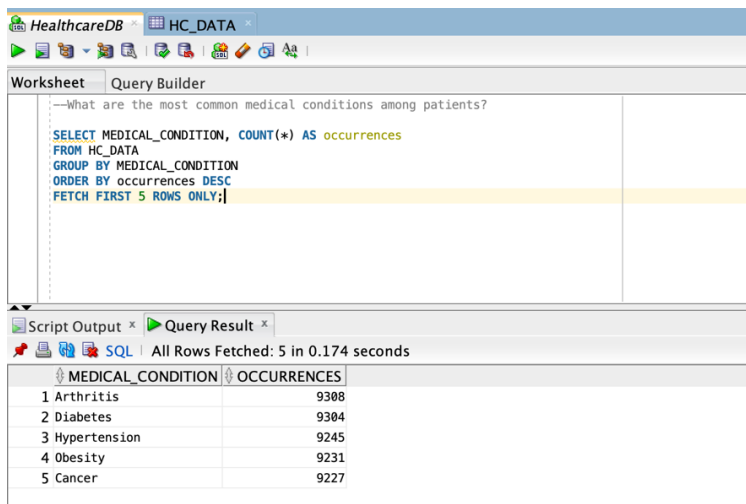


The screenshot shows a database query interface with a tab labeled 'HealthcareDB' and a sub-tab 'HC_DATA'. The query is entered in the 'Query Builder' section. The query is: `SELECT ADMISSION_TYPE, COUNT(*) AS count FROM HC_DATA GROUP BY ADMISSION_TYPE ORDER BY count DESC;`. The 'Script Output' and 'Query Result' tabs are visible at the bottom. The 'Query Result' tab shows the results of the query, which are 3 rows: Elective (18655), Urgent (18576), and Emergency (18269).

ADMISSION_TYPE	COUNT
1 Elective	18655
2 Urgent	18576
3 Emergency	18269

What are the most common medical conditions among patients?

- **Query:** We use `GROUP BY MEDICAL_CONDITION` and `COUNT (*)`, sorting by the highest occurrence.
- **Insight:** Assists in prioritizing specialized treatments and ensuring the availability of necessary medications.



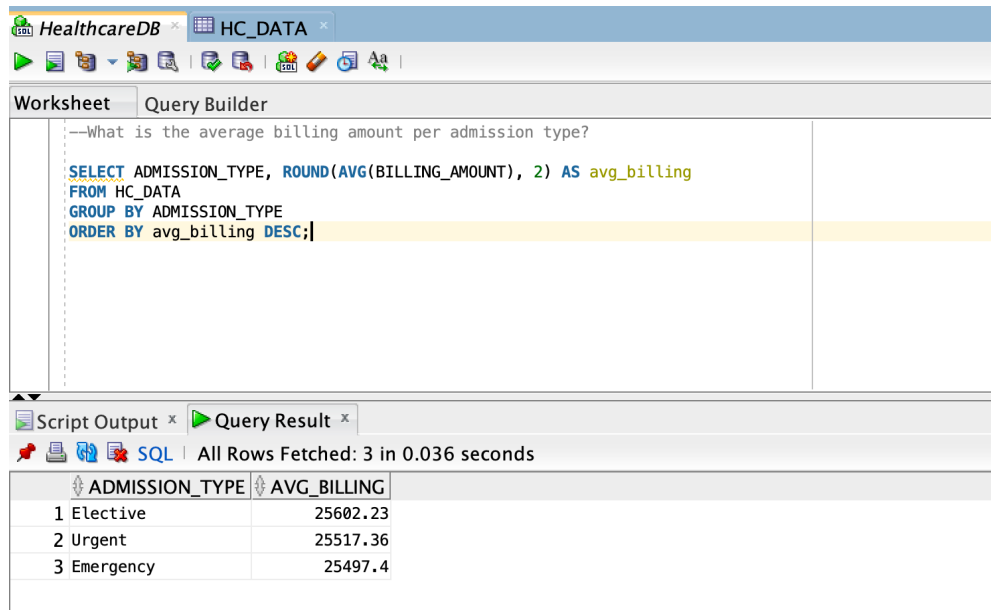
The screenshot shows a database query interface with a tab labeled 'HealthcareDB' and a sub-tab 'HC_DATA'. The query is entered in the 'Query Builder' section. The query is: `SELECT MEDICAL_CONDITION, COUNT(*) AS occurrences FROM HC_DATA GROUP BY MEDICAL_CONDITION ORDER BY occurrences DESC FETCH FIRST 5 ROWS ONLY;`. The 'Script Output' and 'Query Result' tabs are visible at the bottom. The 'Query Result' tab shows the results of the query, which are 5 rows: Arthritis (9308), Diabetes (9304), Hypertension (9245), Obesity (9231), and Cancer (9227).

MEDICAL_CONDITION	OCCURRENCES
1 Arthritis	9308
2 Diabetes	9304
3 Hypertension	9245
4 Obesity	9231
5 Cancer	9227

2. Financial & Insurance Analysis

What is the average billing amount per admission type?

- **Query:** We use `AVG(BILLING_AMOUNT)` to compute the average billing for each admission type.
- **Insight:** Helps in setting pricing policies for different admission types.

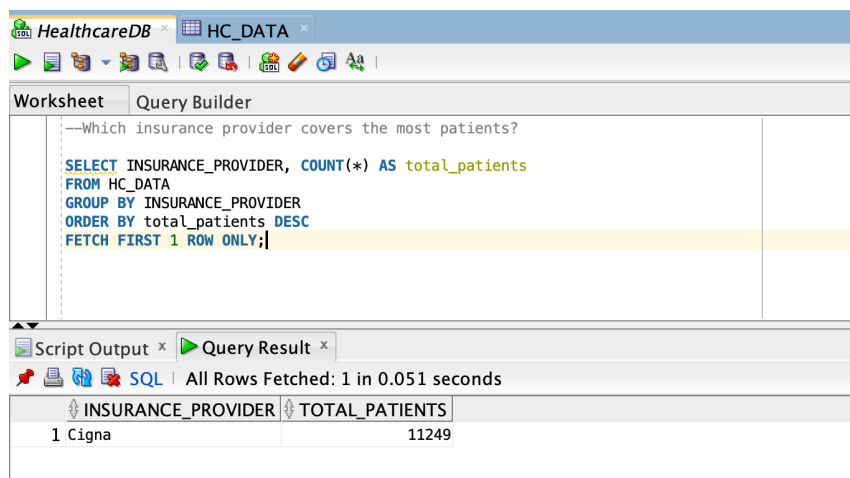


The screenshot shows a database query tool interface. The top bar indicates the database is 'HealthcareDB' and the current table is 'HC_DATA'. The 'Query Builder' tab is active, displaying a SQL query: `SELECT ADMISSION_TYPE, ROUND(AVG(BILLING_AMOUNT), 2) AS avg_billing FROM HC_DATA GROUP BY ADMISSION_TYPE ORDER BY avg_billing DESC;`. Below the query, the 'Query Result' tab shows the results of the query, which are 3 rows. The results are displayed in a table with two columns: 'ADMISSION_TYPE' and 'AVG_BILLING'.

ADMISSION_TYPE	AVG_BILLING
1 Elective	25602.23
2 Urgent	25517.36
3 Emergency	25497.4

Which insurance provider covers the most patients?

- **Query:** Using `GROUP BY INSURANCE_PROVIDER` and `COUNT(*)`, we determine the most used insurance provider.
- **Insight:** Allows hospitals to understand which insurers they work with most frequently for better contract negotiations.

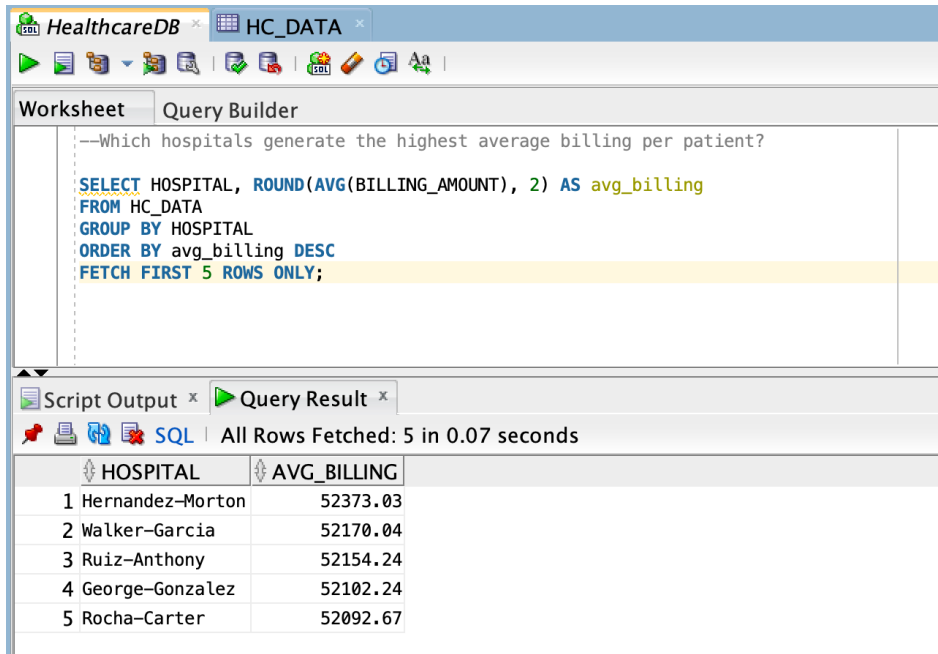


The screenshot shows a database query tool interface. The top bar indicates the database is 'HealthcareDB' and the current table is 'HC_DATA'. The 'Query Builder' tab is active, displaying a SQL query: `SELECT INSURANCE_PROVIDER, COUNT(*) AS total_patients FROM HC_DATA GROUP BY INSURANCE_PROVIDER ORDER BY total_patients DESC FETCH FIRST 1 ROW ONLY;`. Below the query, the 'Query Result' tab shows the results of the query, which are 1 row. The results are displayed in a table with two columns: 'INSURANCE_PROVIDER' and 'TOTAL_PATIENTS'.

INSURANCE_PROVIDER	TOTAL_PATIENTS
1 Cigna	11249

Which hospitals generate the highest average billing per patient?

- **Query:** We use `AVG(BILLING_AMOUNT)` with `GROUP BY HOSPITAL`, sorting results in descending order.
- **Insight:** Helps benchmark hospital revenue performance and financial efficiency.



The screenshot shows a SQL query editor with a query to find the top 5 hospitals by average billing per patient. The query is as follows:

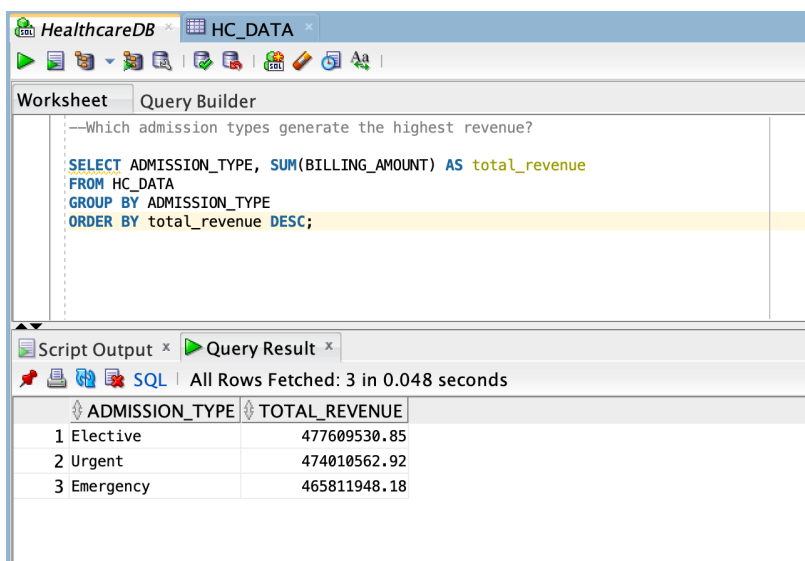
```
--Which hospitals generate the highest average billing per patient?  
  
SELECT HOSPITAL, ROUND(AVG(BILLING_AMOUNT), 2) AS avg_billing  
FROM HC_DATA  
GROUP BY HOSPITAL  
ORDER BY avg_billing DESC  
FETCH FIRST 5 ROWS ONLY;
```

The query result is displayed in a table with 5 rows and 2 columns: HOSPITAL and AVG_BILLING.

HOSPITAL	AVG_BILLING
1 Hernandez-Morton	52373.03
2 Walker-Garcia	52170.04
3 Ruiz-Anthony	52154.24
4 George-Gonzalez	52102.24
5 Rocha-Carter	52092.67

Which admission types generate the highest total revenue?

- **Query:** We use `SUM(BILLING_AMOUNT)` to find the total revenue per admission type.
- **Insight:** Guides hospitals in allocating resources to high-revenue departments.



The screenshot shows a SQL query editor with a query to find the top 3 admission types by total revenue. The query is as follows:

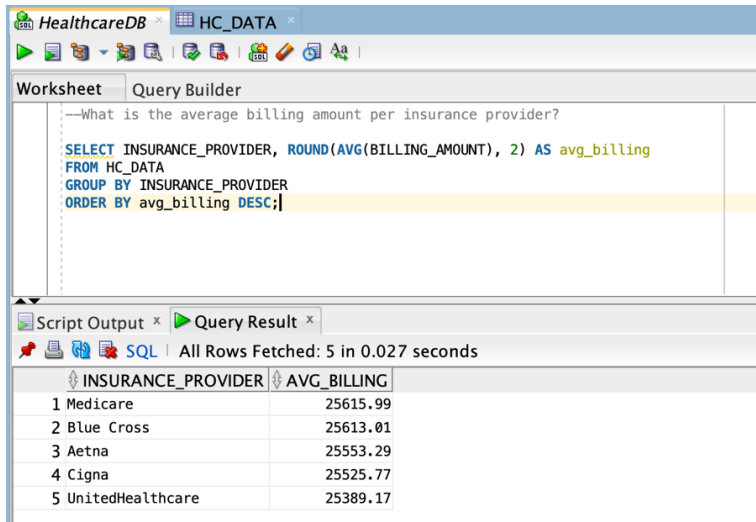
```
--Which admission types generate the highest revenue?  
  
SELECT ADMISSION_TYPE, SUM(BILLING_AMOUNT) AS total_revenue  
FROM HC_DATA  
GROUP BY ADMISSION_TYPE  
ORDER BY total_revenue DESC;
```

The query result is displayed in a table with 3 rows and 2 columns: ADMISSION_TYPE and TOTAL_REVENUE.

ADMISSION_TYPE	TOTAL_REVENUE
1 Elective	477609530.85
2 Urgent	474010562.92
3 Emergency	465811948.18

What is the average billing amount per insurance provider?

- **Query:** We use `AVG(BILLING_AMOUNT)` grouped by `INSURANCE_PROVIDER`.
- **Insight:** Helps in understanding cost variations across different insurance plans.



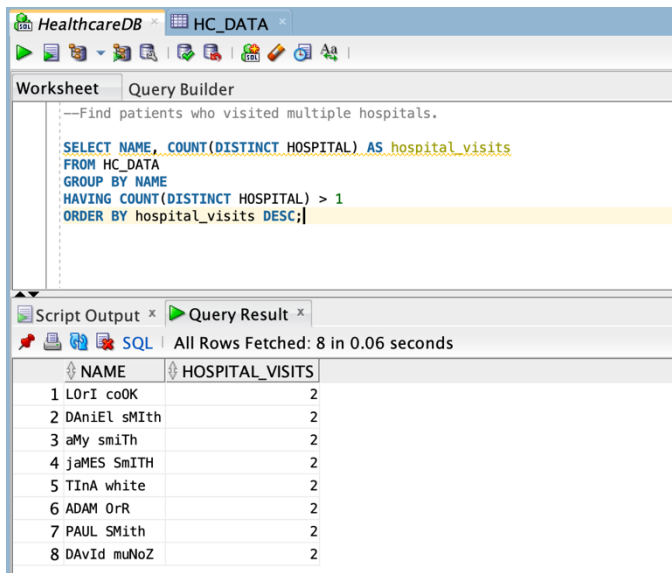
The screenshot shows a database query tool interface. The top bar has tabs for 'HealthcareDB' and 'HC_DATA'. Below the toolbar, the 'Query Builder' tab is active, displaying a query: `SELECT INSURANCE_PROVIDER, ROUND(AVG(BILLING_AMOUNT), 2) AS avg_billing FROM HC_DATA GROUP BY INSURANCE_PROVIDER ORDER BY avg_billing DESC;`. Below the query, the 'Query Result' tab shows the results of the query, which are 5 rows. The columns are 'INSURANCE_PROVIDER' and 'AVG_BILLING'.

INSURANCE_PROVIDER	AVG_BILLING
1 Medicare	25615.99
2 Blue Cross	25613.01
3 Aetna	25553.29
4 Cigna	25525.77
5 UnitedHealthcare	25389.17

3. Advanced Healthcare Insights

Find patients who visited multiple hospitals.

- **Query:** We use `COUNT(DISTINCT HOSPITAL)` with `HAVING COUNT(DISTINCT HOSPITAL) > 1` to identify such patients.
- **Insight:** Helps track patient loyalty and referral trends across hospitals.



The screenshot shows a database query tool interface. The top bar has tabs for 'HealthcareDB' and 'HC_DATA'. Below the toolbar, the 'Query Builder' tab is active, displaying a query: `SELECT NAME, COUNT(DISTINCT HOSPITAL) AS hospital_visits FROM HC_DATA GROUP BY NAME HAVING COUNT(DISTINCT HOSPITAL) > 1 ORDER BY hospital_visits DESC;`. Below the query, the 'Query Result' tab shows the results of the query, which are 8 rows. The columns are 'NAME' and 'HOSPITAL_VISITS'.

NAME	HOSPITAL_VISITS
1 LOrI coOK	2
2 DAniEl sMIth	2
3 aMy smIth	2
4 JaMES SMITH	2
5 TInA white	2
6 ADAM OrR	2
7 PAUL SMith	2
8 DAVID muNoZ	2

What is the readmission rate per doctor?

- **Query:** We perform a **self-join** on HC_DATA to track patients who were readmitted within 30 days, grouping by doctor.
- **Insight:** Evaluate doctor performance and effectiveness of treatment plans.

The screenshot shows the HealthcareDB interface with a query builder. The query is designed to find the doctor with the highest patient readmission rate within 30 days of discharge. The query uses a self-join on the HC_DATA table, grouping by doctor and ordering by readmission count in descending order, fetching the top 5 results.

```
--Which doctor has the highest patient readmission rate?  
  
SELECT A.DOCTOR, COUNT(DISTINCT A.NAME) AS readmission_count  
FROM HC_DATA A  
JOIN HC_DATA B  
ON A.NAME = B.NAME  
AND B.DATE_OF_ADMISSION > A.DISCHARGE_DATE  
AND B.DATE_OF_ADMISSION - A.DISCHARGE_DATE <= 30  
GROUP BY A.DOCTOR  
ORDER BY readmission_count DESC  
FETCH FIRST 5 ROWS ONLY;
```

The query result shows one row:

DOCTOR	READMISSION_COUNT
1 Craig Davis	1

Which medications are prescribed most often for high-billing cases?

- **Query:** We use GROUP BY MEDICATION, count prescriptions, and calculate AVG (BILLING_AMOUNT).
- **Insight:** Helps in understanding which medications are driving high treatment costs.

The screenshot shows the HealthcareDB interface with a query builder. The query is designed to find the medications prescribed most often for high-billing cases. The query uses GROUP BY MEDICATION, counting prescriptions and calculating the average billing amount, ordering by average cost in descending order, and fetching the top 5 results.

```
--Which medications are prescribed most often for high-billing cases?  
  
SELECT MEDICATION, COUNT(*) AS prescription_count, ...  
ROUND(AVG(BILLING_AMOUNT), 2) AS avg_cost  
FROM HC_DATA  
GROUP BY MEDICATION  
ORDER BY avg_cost DESC  
FETCH FIRST 5 ROWS ONLY;
```

The query result shows five rows:

MEDICATION	PRESCRIPTION_COUNT	AVG_COST
1 Ibuprofen	11127	25735.58
2 Aspirin	11094	25594.26
3 Paracetamol	11071	25533.47
4 Penicillin	11068	25490.92
5 Lipitor	11140	25342.47

Which doctor has the highest patient readmission rate?

- **Query:** Using `JOIN` and `COUNT(DISTINCT NAME)`, we identify doctors with the highest number of returning patients.
- **Insight:** Assists hospitals in monitoring the **quality of care**.

The screenshot shows a database query tool with a 'Query Builder' tab. The query is: `--Which doctor has the highest patient readmission rate?`
`SELECT A.DOCTOR, COUNT(DISTINCT A.NAME) AS readmission_count`
`FROM HC_DATA A`
`JOIN HC_DATA B`
`ON A.NAME = B.NAME`
`AND B.DATE_OF_ADMISSION > A.DISCHARGE_DATE`
`AND B.DATE_OF_ADMISSION - A.DISCHARGE_DATE <= 30`
`GROUP BY A.DOCTOR`
`ORDER BY readmission_count DESC`
`FETCH FIRST 5 ROWS ONLY;`

The 'Query Result' tab shows the results:

DOCTOR	READMISSION_COUNT
1 Craig Davis	1

What is the average doctor workload (patients per doctor per month)?

- **Query:** We use `EXTRACT(MONTH FROM DATE_OF_ADMISSION)`, `GROUP BY DOCTOR`, and `COUNT(*)`.
- **Insight:** Supports hospital scheduling and workload management for medical professionals

The screenshot shows a database query tool with a 'Query Builder' tab. The query is: `--What is the average doctor workload (patients per doctor per month)?`
`SELECT DOCTOR,`
`EXTRACT(MONTH FROM DATE_OF_ADMISSION) AS month,`
`COUNT(*) AS patients_treated`
`FROM HC_DATA`
`GROUP BY DOCTOR, EXTRACT(MONTH FROM DATE_OF_ADMISSION)`
`ORDER BY month DESC, patients_treated DESC;`

The 'Query Result' tab shows the results:

DOCTOR	MONTH	PATIENTS_TREATED
1 Brian Rodriguez	12	4
2 John Jackson	12	4
3 Michael Johnson	12	3
4 William Moore	12	3
5 Patricia Wilson	12	3
6 Michael Brown	12	3
7 Annette Smith	12	3

Conclusion

This project successfully demonstrates how **SQL can be utilized to analyze healthcare data**, providing **valuable insights into patient demographics, hospital efficiency, financial management, and medical trends**. By leveraging SQL queries, we uncovered key trends such as:

- The most common medical conditions requiring hospital attention.
- The distribution of admissions across different categories (Emergency, Urgent, Elective).
- Hospital revenue generation patterns based on admission types and insurance providers.
- The impact of doctor performance on patient readmission rates.
- The relationship between prescribed medications and billing costs.

Key Takeaways:

- Data-driven decisions can optimize hospital operations and resource allocation.
- Financial analysis of billing and insurance data can help maximize hospital revenue and improve pricing strategies.
- Tracking patient behavior (readmissions, hospital switching) aids in improving patient care