# Healthcare analysis report

## Dataset Overview

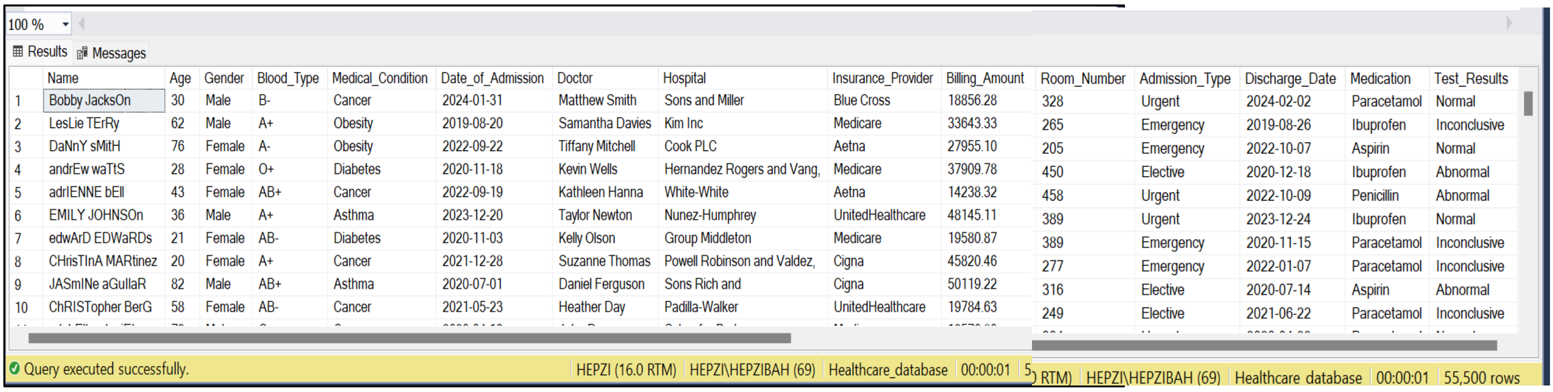
The healthcare_dataset contains structured patient-level data collected from one or more hospitals. It includes key information such as patient demographics, medical conditions, medications, admission and discharge dates, assigned doctors, billing amounts, and insurance details. Each row represents a single patient visit or encounter. This dataset enables analysis of hospital operations, including revenue trends, doctor performance, patient readmissions, and insurance coverage. It also supports demographic and clinical pattern analysis. The dataset is suitable for data cleaning, reporting, and dashboarding tasks, making it ideal for healthcare analytics and decision-making in real-world scenarios.

## Purpose of Analysis

This dataset will be used to:

- Identify trends in patient admissions and medical conditions
- Analyze hospital revenue and billing accuracy
- Assess performance of doctors and insurance providers
- Understand demographic patterns like age and gender distribution
- Detect and clean data quality issues (e.g., missing or invalid values)

## Previewing the Healthcare Dataset



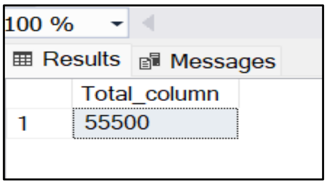
| | Name | Age | Gender | Blood_Type | Medical_Condition | Date_of_Admission | Doctor | Hospital | Insurance_Provider | Billing_Amount | Room_Number | Admission_Type | Discharge_Date | Medication | Test_Results |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Bobby JacksOn | 30 | Male | B- | Cancer | 2024-01-31 | Matthew Smith | Sons and Miller | Blue Cross | 18856.28 | 328 | Urgent | 2024-02-02 | Paracetamol | Normal |
| 2 | LesLie TErRy | 62 | Male | A+ | Obesity | 2019-08-20 | Samantha Davies | Kim Inc | Medicare | 33643.33 | 265 | Emergency | 2019-08-26 | Ibuprofen | Inconclusive |
| 3 | DaNnY sMitH | 76 | Female | A- | Obesity | 2022-09-22 | Tiffany Mitchell | Cook PLC | Aetna | 27955.10 | 205 | Emergency | 2022-10-07 | Aspirin | Normal |
| 4 | andrEw waTtS | 28 | Female | O+ | Diabetes | 2020-11-18 | Kevin Wells | Hernandez Rogers and Vang, | Medicare | 37909.78 | 450 | Elective | 2020-12-18 | Ibuprofen | Abnormal |
| 5 | adrIENNE bEll | 43 | Female | AB+ | Cancer | 2022-09-19 | Kathleen Hanna | White-White | Aetna | 14238.32 | 458 | Urgent | 2022-10-09 | Penicillin | Abnormal |
| 6 | EMILY JOHNSOn | 36 | Male | A+ | Asthma | 2023-12-20 | Taylor Newton | Nunez-Humphrey | UnitedHealthcare | 48145.11 | 389 | Urgent | 2023-12-24 | Ibuprofen | Normal |
| 7 | edwArD EDWaRDs | 21 | Female | AB- | Diabetes | 2020-11-03 | Kelly Olson | Group Middleton | Medicare | 19580.87 | 389 | Emergency | 2020-11-15 | Paracetamol | Inconclusive |
| 8 | CHrisTInA MARtinez | 20 | Female | A+ | Cancer | 2021-12-28 | Suzanne Thomas | Powell Robinson and Valdez, | Cigna | 45820.46 | 277 | Emergency | 2022-01-07 | Paracetamol | Inconclusive |
| 9 | JASmINe aGullaR | 82 | Male | AB+ | Asthma | 2020-07-01 | Daniel Ferguson | Sons Rich and | Cigna | 50119.22 | 316 | Elective | 2020-07-14 | Aspirin | Abnormal |
| 10 | ChRISTopher BerG | 58 | Female | AB- | Cancer | 2021-05-23 | Heather Day | Padilla-Walker | UnitedHealthcare | 19784.63 | 249 | Elective | 2021-06-22 | Paracetamol | Inconclusive |

Query executed successfully. | HEPZI (16.0 RTM) | HEPZI\HEPZIBAH (69) | Healthcare_database | 00:00:01 | 5

) RTM) | HEPZI\HEPZIBAH (69) | Healthcare database | 00:00:01 | 55,500 rows

DATA CLEANING SUMMARY

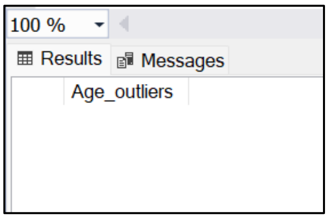
- ➢ Total Number of Columns:



| | Total_column |
|---|---|
| 1 | 55500 |

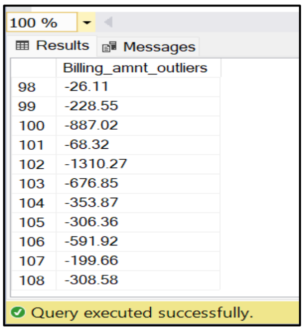
There are 55,500 rows in the healthcare dataset. It is a relatively large dataset.

- ➢ Checking Outliers of Age And Billing Amounts:



| | Age_outliers |
|---|---|

Age:
Fortunately, the Age column contains no significant outliers, indicating that the data is consistent and within an expected range for patient demographics.

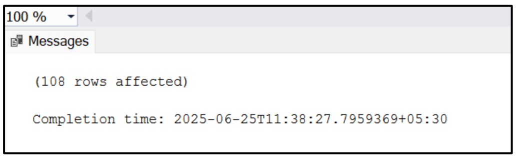


| | Billing_amnt_outliers |
|---|---|
| 98 | -26.11 |
| 99 | -228.55 |
| 100 | -887.02 |
| 101 | -68.32 |
| 102 | -1310.27 |
| 103 | -676.85 |
| 104 | -353.87 |
| 105 | -306.36 |
| 106 | -591.92 |
| 107 | -199.66 |
| 108 | -308.58 |

Query executed successfully.

Billing Amounts:
The Billing Amount column contains multiple negative values, with 108 entries identified as outliers that may indicate billing errors or data quality issues.

# Healthcare analysis report



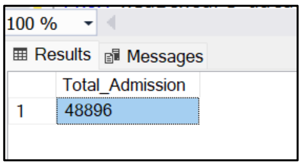
```
(108 rows affected)

Completion time: 2025-06-25T11:38:27.7959369+05:30
```

All negative billing entries were replaced with the average billing amount to ensure consistency in financial metrics.

- All missing (NULL) values have been identified and handled to ensure data completeness.
- Inconsistent values in Admission_Type were unified (e.g., 'ER', 'Emergency Room' → 'Emergency').
- Missing values in the Medication column were replaced with 'Unknown'.
- Records where Discharge_Date occurs before Date_of_Admission were flagged as invalid.
- Entries with Date_of_Admission in the future were detected and marked for review.
- A new PatientID column was created using the DENSE_RANK() function to uniquely identify patients.

## Key Metrics & Analysis

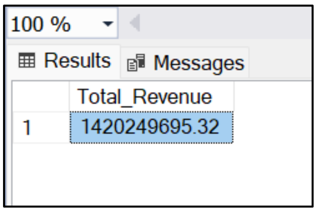
### 1.Total number of Admission:



| | Total_Admission |
|---|---|
| 1 | 48896 |

Counts the number of unique patients (PatientID) to calculate the total number of admissions handled by the hospital.

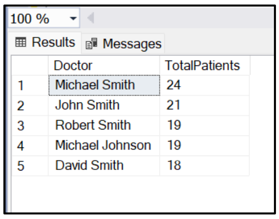
### 2.Total revenue of the Hospital:



| | Total_Revenue |
|---|---|
| 1 | 1420249695.32 |

Calculates the overall billing revenue by summing the Billing_Amount column across all records.

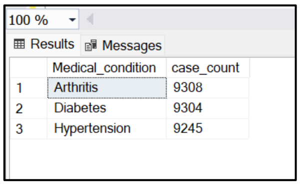
### 3. Top 5 Doctors by Number of Patients :



| | Doctor | TotalPatients |
|---|---|---|
| 1 | Michael Smith | 24 |
| 2 | John Smith | 21 |
| 3 | Robert Smith | 19 |
| 4 | Michael Johnson | 19 |
| 5 | David Smith | 18 |

Identifies the five doctors who treated the highest number of unique patients

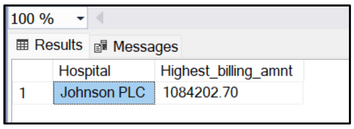
### 4. What are the most common medical conditions?



| | Medical_condition | case_count |
|---|---|---|
| 1 | Arthritis | 9308 |
| 2 | Diabetes | 9304 |
| 3 | Hypertension | 9245 |

Lists the top three frequently occurring medical conditions among all patients.

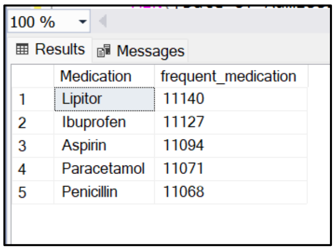
### 5. Which hospital handles the highest billing amount?



| | Hospital | Highest_billing_amnt |
|---|---|---|
| 1 | Johnson PLC | 1084202.70 |

Determines which hospital generated the most revenue based on total billing amounts.

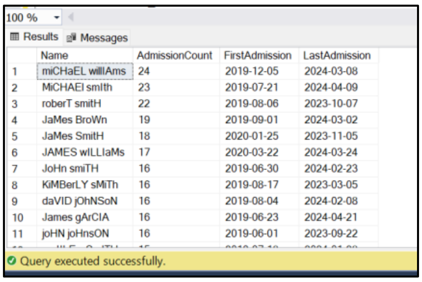
### 6.Which medications are most frequently used?



| | Medication | frequent_medication |
|---|---|---|
| 1 | Lipitor | 11140 |
| 2 | Ibuprofen | 11127 |
| 3 | Aspirin | 11094 |
| 4 | Paracetamol | 11071 |
| 5 | Penicillin | 11068 |

Displays the medications prescribed most often, ranked by count of occurrences.

## 7.Patient Readmission Analysis:

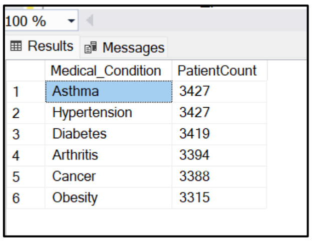


| | Name | AdmissionCount | FirstAdmission | LastAdmission |
|---|---|---|---|---|
| 1 | miCHaEL willAms | 24 | 2019-12-05 | 2024-03-08 |
| 2 | MiCHAEl smIth | 23 | 2019-07-21 | 2024-04-09 |
| 3 | roberT smitH | 22 | 2019-08-06 | 2023-10-07 |
| 4 | JaMes BroWn | 19 | 2019-09-01 | 2024-03-02 |
| 5 | JaMes SmitH | 18 | 2020-01-25 | 2023-11-05 |
| 6 | JAMES wILLIaMs | 17 | 2020-03-22 | 2024-03-24 |
| 7 | JoHn smiTH | 16 | 2019-06-30 | 2024-02-23 |
| 8 | KiMBerLY sMiTh | 16 | 2019-08-17 | 2023-03-05 |
| 9 | daVID jOhNSoN | 16 | 2019-08-04 | 2024-02-08 |
| 10 | James gArCIA | 16 | 2019-06-23 | 2024-04-21 |
| 11 | joHN joHnsON | 16 | 2019-06-01 | 2023-09-22 |

Query executed successfully.

Highlights patients with multiple admissions and shows their first and last admission dates.

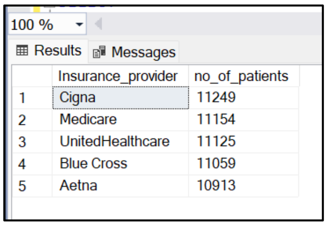
## 8. Most Common Conditions (for Age > 60)



| | Medical_Condition | PatientCount |
|---|---|---|
| 1 | Asthma | 3427 |
| 2 | Hypertension | 3427 |
| 3 | Diabetes | 3419 |
| 4 | Arthritis | 3394 |
| 5 | Cancer | 3388 |
| 6 | Obesity | 3315 |

Lists medical conditions most commonly reported among elderly patients above 60 years of age.

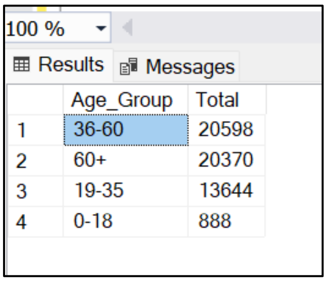
## 9.Top Insurance Providers by Patient Volume



| | Insurance_provider | no_of_patients |
|---|---|---|
| 1 | Cigna | 11249 |
| 2 | Medicare | 11154 |
| 3 | UnitedHealthcare | 11125 |
| 4 | Blue Cross | 11059 |
| 5 | Aetna | 10913 |

Ranks insurance companies by the number of patients they cover in the dataset.

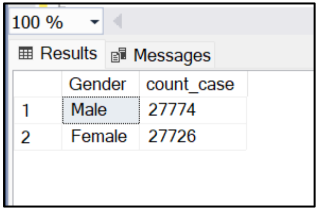
## 10.Age group distribution:



| | Age_Group | Total |
|---|---|---|
| 1 | 36-60 | 20598 |
| 2 | 60+ | 20370 |
| 3 | 19-35 | 13644 |
| 4 | 0-18 | 888 |

Categorizes patients into age groups and counts how many fall into each, showing the hospital's demographic spread.

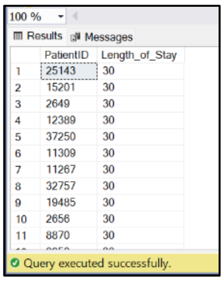
## 11.Do males or females get admitted more often?



| | Gender | count_case |
|---|---|---|
| 1 | Male | 27774 |
| 2 | Female | 27726 |

Analyzes whether more males or females are being admitted by counting gender-wise admissions.

## 12.Length of stay Analysis :



| | PatientID | Length_of_Stay |
|---|---|---|
| 1 | 25143 | 30 |
| 2 | 15201 | 30 |
| 3 | 2649 | 30 |
| 4 | 12389 | 30 |
| 5 | 37250 | 30 |
| 6 | 11309 | 30 |
| 7 | 11267 | 30 |
| 8 | 32757 | 30 |
| 9 | 19485 | 30 |
| 10 | 2656 | 30 |
| 11 | 8870 | 30 |

Query executed successfully.

Calculates the number of days each patient stayed in the hospital using admission and discharge dates

## Conclusion:

The healthcare dataset analysis uncovered key patterns in admissions, billing, patient demographics, and medical conditions. After thorough data cleaning, accurate insights were derived on hospital revenue, doctor performance, and patient trends.

This project demonstrates the application of SQL for real-world healthcare analytics, enabling hospitals to improve operational efficiency, enhance quality of care, and make informed decisions using data.