Hospital Charge Analysis
Non-Technical Stakeholder Report
Written by Tolulope Olarewaju
16th June 2025

Table of Contents

U.S. Hospital Charges Analysis

This project presents an end-to-end data analysis on U.S. inpatient hospital charges using SQL, Excel, and Power BI. It is featured as a case study on my portfolio and demonstrates my skills in querying, business reasoning, and data visualization relevant to healthcare analytics.

Project Summary:

Hospital charges for the same diagnosis can vary widely across U.S. states and providers. This project analyzes those variations to support transparency, cost control, and smarter healthcare decisions.

- Dataset Size: 163,065 rows
- Source: Kaggle – Inpatient Hospital Charges Dataset
- Skills Demonstrated: SQL querying, data cleaning, Power BI dashboarding, healthcare analytics

Project Objectives:

- Identify hospitals and states with unusually high or low inpatient charges.
- Analyze the gap between hospital charges and Medicare payments.
- Highlight the most common diagnoses by discharge volume.
- Present insights through interactive dashboards for non-technical audiences.

Tools and Technologies Used:

- MySQL — SQL queries for data exploration and analysis
- Excel — Initial data cleaning (e.g., removing dollar signs)
- Power BI — Interactive dashboard creation

Data Cleaning & Preparation:

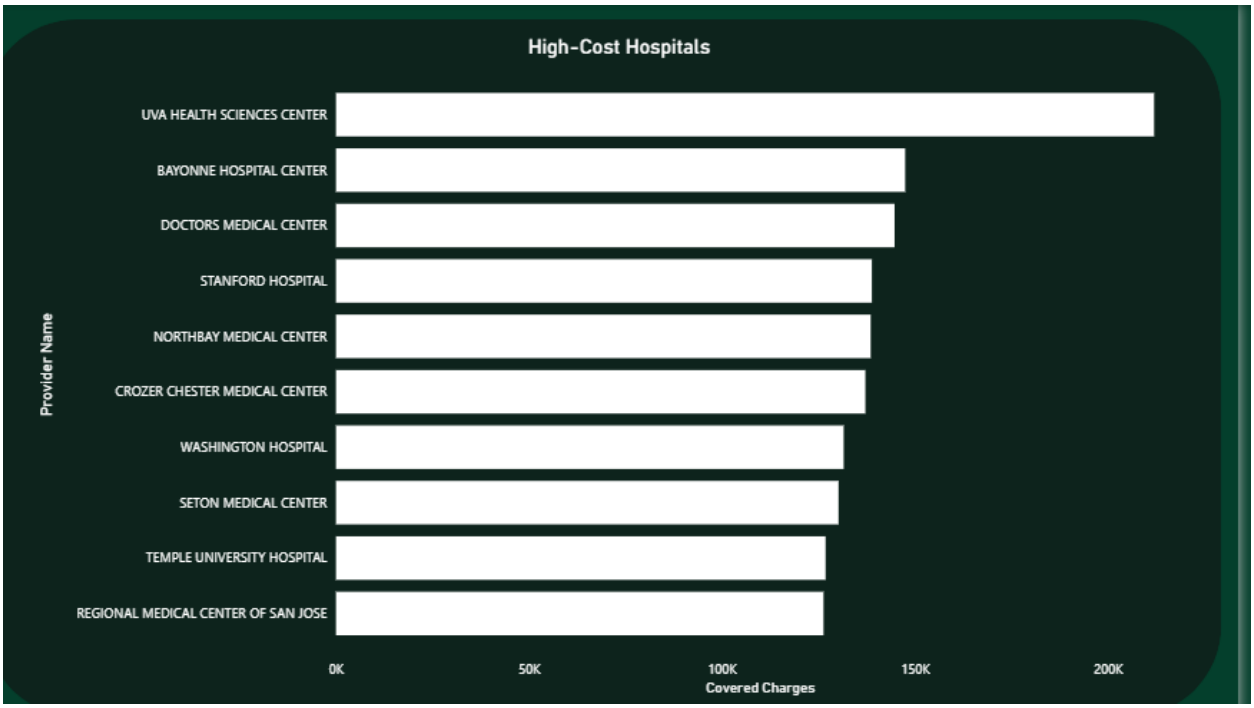
- Removed $ symbols from columns:
    - average_covered_charges
    - average_total_payments
    - average_medicare_payments
- Verified no null or missing key fields (drg_definition, provider_id).

- Loaded data into MySQL Workbench for querying.

SQL Analysis Queries:

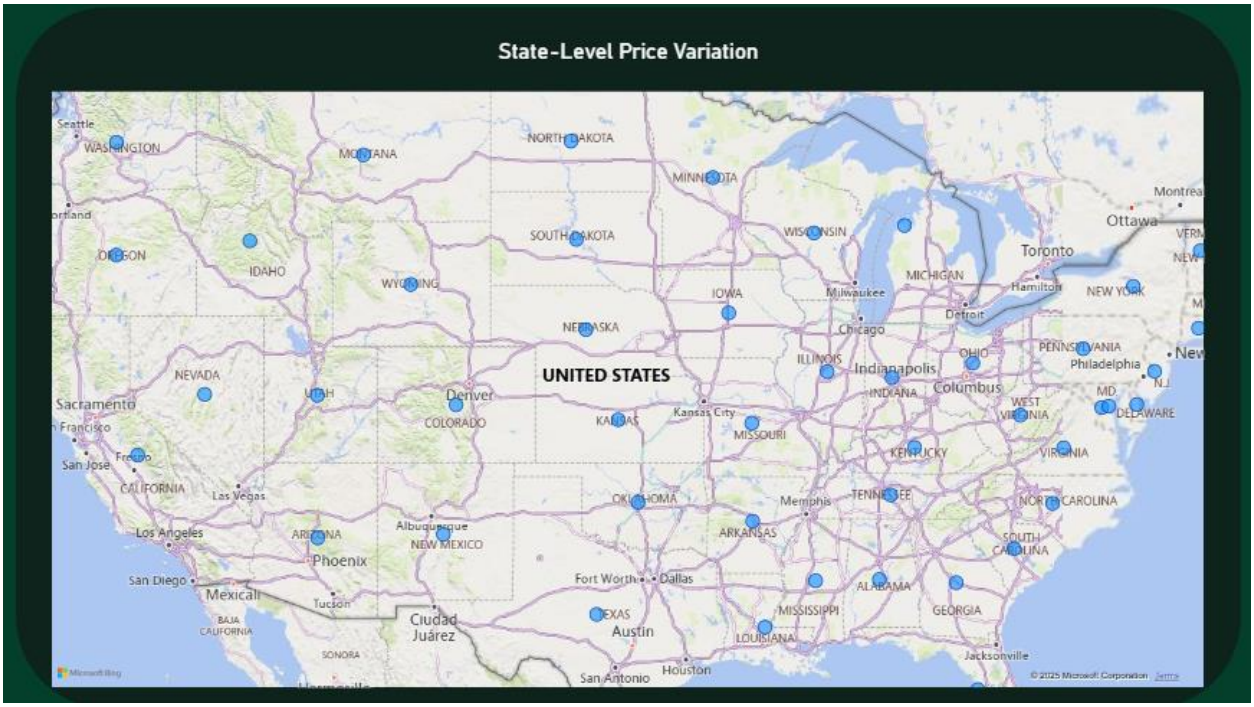
1 Identifying High-Cost Hospitals

```
SELECT provider_name, provider_city, provider_state,
ROUND(AVG(average_covered_charges), 2) AS avg_covered_charges
FROM inpatient_charges
GROUP BY provider_name, provider_city, provider_state
ORDER BY avg_covered_charges DESC
LIMIT 10;
```



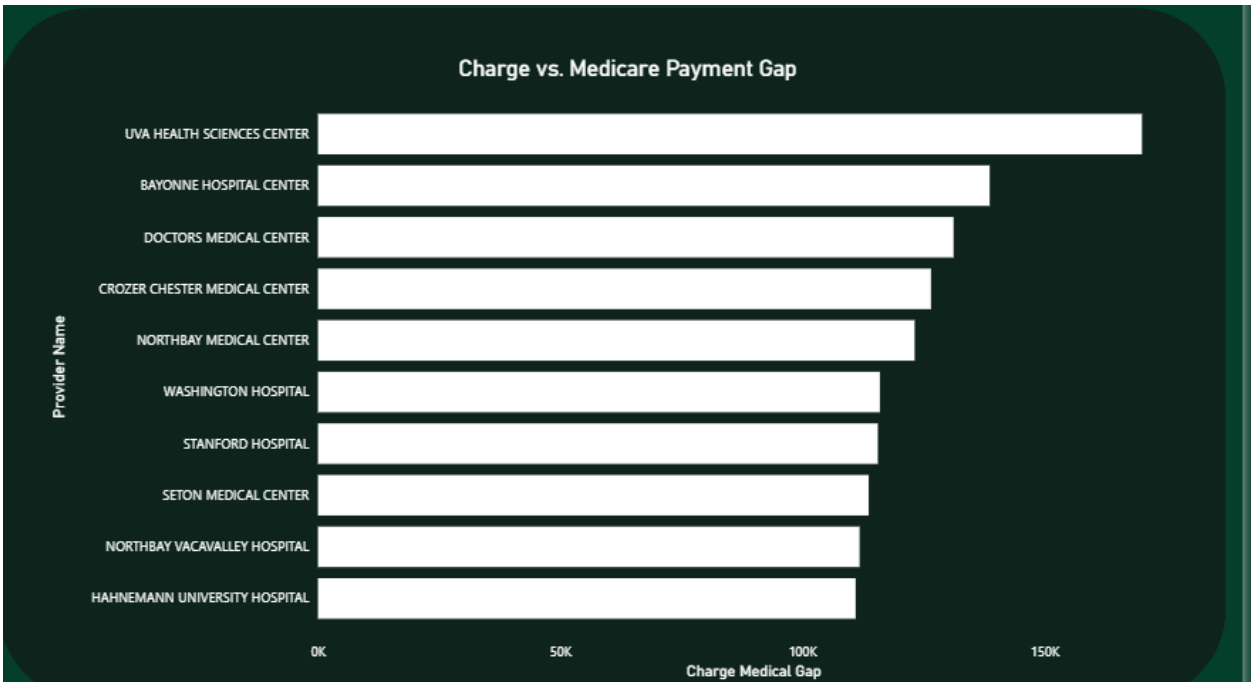
2 Understanding Pricing Differences Across States

```
SELECT provider_state,
ROUND(AVG(average_covered_charges), 2) AS avg_state_charges
FROM inpatient_charges
GROUP BY provider_state
ORDER BY avg_state_charges DESC;
```



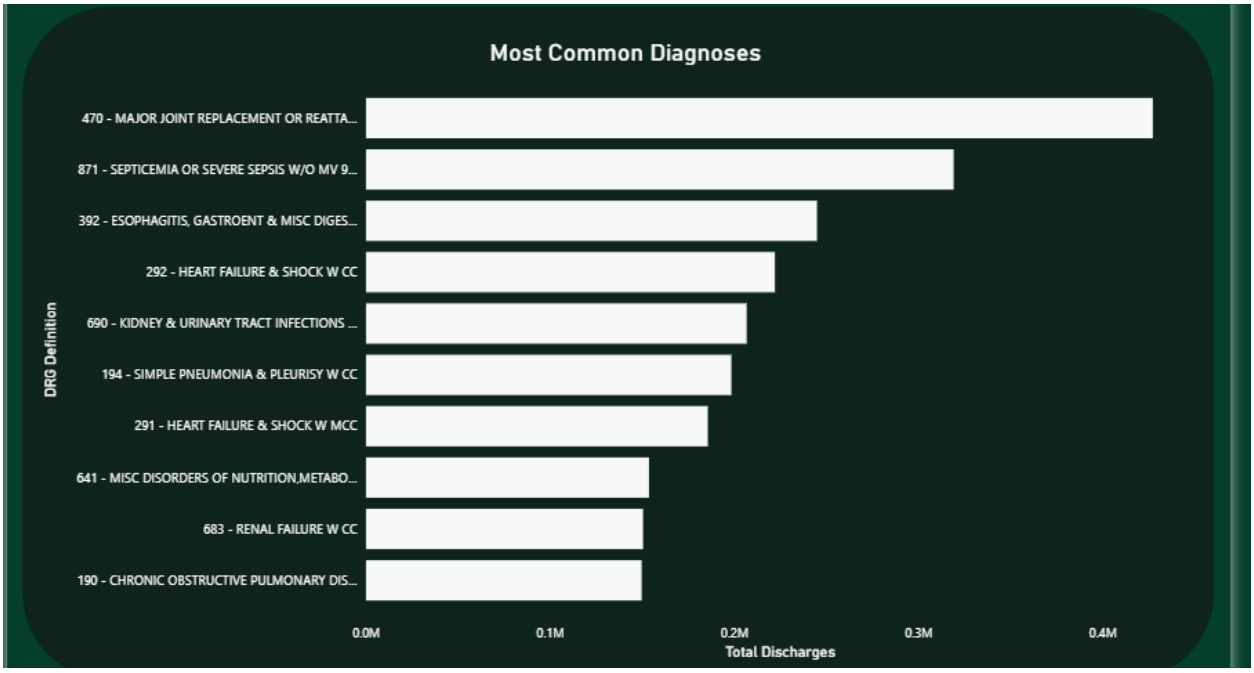
3 Analyzing Profit Gaps Between Hospitals

```
SELECT provider_name,
ROUND(AVG(average_covered_charges - average_medicare_payments), 2) AS
charge_medicare_gap
FROM inpatient_charges
GROUP BY provider_name
ORDER BY charge_medicare_gap DESC
LIMIT 10;
```



4 Finding the Most Common Diagnoses by Volume

```
SELECT drg_definition,
SUM(total_discharges) AS total_discharges
FROM inpatient_charges
GROUP BY drg_definition
ORDER BY total_discharges DESC
LIMIT 10;
```



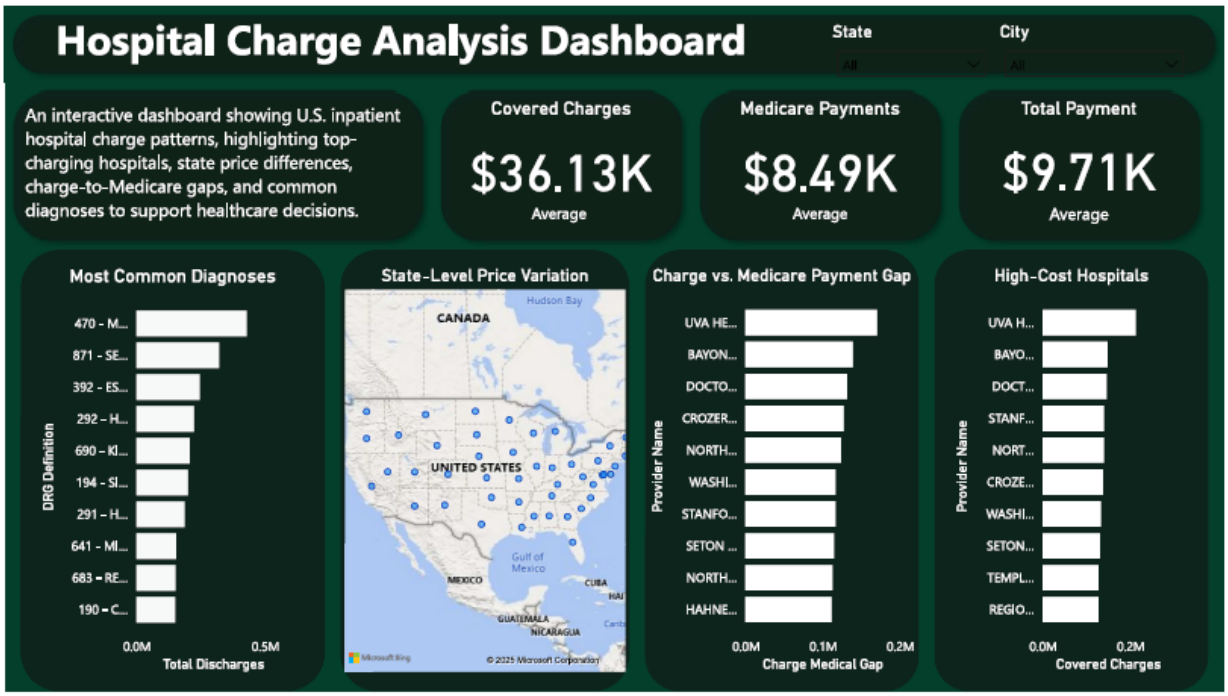
Power BI Dashboard Overview:

Dashboard Pages:

- High-Cost Hospitals
- State Price Variations
- Medicare Payment Gaps
- Common Diagnoses by Volume

Key Visuals:

- Bar Charts
- Map Visualizations
- Card KPIs



Insights:

- California and New Jersey have the highest average hospital charges.
- Some hospitals charge over $97,000 more than Medicare reimburses.
- Joint replacements and septicemia are the most common inpatient diagnoses.
- 163,065 rows were processed, showing robust SQL and data handling capacity.

Challenges & Solutions:

Challenge	Solution
Dollar signs in numeric columns	Cleaned using Excel before SQL import
Slow data import into MySQL	Reduced file size and checked column formats
Communicating technical queries to non-technical users	Used Power BI dashboards with simple visuals

Recommendations:

- Monitor hospitals with large charge-to-Medicare gaps for potential overpricing.
- Improve healthcare cost transparency at the state and provider level.
- Prioritize resource planning around the most common diagnoses by discharge volume.

Dataset Source:

- Kaggle: Inpatient Hospital Charges Dataset
- Original Source: data.gov