# **HealthData IQ – Hospital Insights & Patient Analytics**

## **Project Overview**

This project analyzes hospital performance and patient feedback across the United States using data from the Centers for Medicare & Medicaid Services (CMS). The analysis aims to uncover trends in quality, satisfaction, emergency services, and hospital types to support data-driven healthcare insights.

## **Dataset Summary**

* **Source:**<https://www.kaggle.com/datasets/CMS/hospital-general-information>
* **Data Columns:** Hospital name, type, ownership, ratings, emergency services, mortality, readmission, safety, patient survey results, location.
* **Primary Focus:** Hospital types, ratings, emergency service availability, quality metrics, and ownership.

## **Data Cleaning**

* Removed rows with missing "County Name"
* Dropped "Meets criteria for meaningful use of EHRs" and footnote columns due to high missing values
* Converted key comparison columns (e.g., "Above the national average") to ordinal numeric scores
* Converted "Hospital overall rating" to numeric for statistical use
* Standardized column names and handled inconsistent values (e.g., "Not Available")

## **Exploratory Data Analysis (EDA)**

### **1. Hospital Type and Ownership**

* The most common hospital type is "Acute Care Hospitals," indicating a skewed distribution.
* "Voluntary non-profit - Private" and "Proprietary" are the most frequent ownership types.

### **2. Hospital Ratings by Ownership**

* Most ownership types show a median overall rating around 3.
* "Physician"-owned hospitals show a slightly higher median.
* The presence of "Not Available" ratings affects comparison reliability.

### **3. Emergency Services Availability**

* Among the top 10 states by hospital count, a high proportion of hospitals offer emergency services.
* Emergency services are widely available in large states such as TX and CA.

### **4. Patient Experience**

* Most hospitals are rated "Same as the national average" for patient experience.

### **5. Correlation Analysis**

* Correlation between 'Hospital overall rating', 'ZIP Code', and 'Phone Number' is very weak.
* A heatmap of correlations between overall rating and quality metric scores (mortality, readmission, safety, etc.) shows weak relationships, indicating these metrics measure distinct aspects.

### **6. Mortality Comparison**

* Many hospitals report mortality rates as "Same as the national average."
* Significant counts also for "Below" and "Above" average.

### **7. Chi-Square Test: Hospital Type vs Patient Experience**

* **Chi-Square Statistic:** 1831.28
* **P-Value:** 0.0000
* **Conclusion:** Significant association between hospital type and patient experience ratings.

### **8. Boxplot Analysis**

* Boxplot visualizations show the spread and median of overall hospital ratings across ownership types.

## **Visualizations**

* Histogram: Rating distributions
* Bar Chart: Ratings by ownership
* Pie Chart: Emergency service availability
* Correlation Heatmap: Ratings vs quality metrics
* Boxplot: Ratings by ownership type
* Interactive Dashboard (Plotly):
  + Hospital Type Count
  + Hospital Ownership Distribution
  + Hospital Ratings by Ownership
  + Emergency Services Availability (Top 10 States)
  + Patient Experience Comparison Distribution

## **Dashboard Implementation**

* Built **Plotly**
* Features:
  + State and ownership filters
  + Dynamic charts and interactive visuals
  + Combined subplot layout for overview

## **Key Insights**

* Acute Care and Non-Profit hospitals dominate the dataset.
* Most hospitals have average ratings around 3.
* Emergency services are widely accessible in states with more hospitals.
* Weak correlations among quality metrics suggest independent performance indicators.
* Patient experience significantly varies by hospital type.

## **Limitations**

* High volume of missing data in quality metrics and rating fields.
* Incomplete data limits statistical analysis robustness.

## **Recommendations**

* Promote full reporting of quality metrics and overall ratings.
* Use quality metrics as complementary indicators of performance.
* Improve data completeness in future datasets.

## **Tools Used**

* **Pandas, NumPy, Seaborn, Matplotlib, Plotly**
* **SciPy** for Chi-Square and correlation analysis

## **Project Members**

* Data Cleaning & Analysis:
* Visualization & Reporting:
* Dashboard Development: