# HEALTHCARE ANALYSIS REPORT

**Prepared By: Benjamin Onyatta**

**Date: 4/27/2025**

# Abstract

This report presents a comprehensive analysis of patient care patterns, hospital performance, and treatment costs using a dataset of 55,500 unique patient records from 10 major hospitals across the United States. The dataset includes key variables such as patient demographics, medical conditions, admission types, billing amounts, medications, test results, and hospital locations. Through data visualization in Power BI and exploration in Excel, the analysis addresses critical questions around the most common patient profiles, frequently diagnosed conditions, hospital stay durations, cost variations, medication usage, insurance coverage, and regional differences in healthcare delivery.

The findings reveal notable trends, such as demographic groups with higher admission rates, conditions associated with longer hospital stays and higher costs, discrepancies in treatment costs across hospitals and insurers, and patterns in test outcomes and medication prescriptions. Regional differences in healthcare quality and billing practices are also highlighted. These insights aim to guide healthcare stakeholders in making informed, data-driven decisions to enhance patient outcomes, optimize resource allocation, reduce unnecessary costs, and promote more equitable healthcare practices across institutions and regions.


# Introduction

## Context

The healthcare industry faces increasing demands to deliver high-quality patient care while managing rising treatment costs and addressing disparities in health outcomes. As hospitals across the United States serve diverse populations with varying needs, there is a growing need for data-driven insights to guide clinical, operational, and financial decisions.

This project analyzes a comprehensive dataset containing 55,500 patient records from 10 major hospitals, capturing information about patient demographics, medical conditions, hospital admissions, treatment costs, medications, test outcomes, insurance providers, and geographic locations. The analysis is designed to uncover patterns and trends in patient care, treatment practices, cost structures, and outcomes across different hospitals and regions.

By leveraging Power BI for interactive data visualization, this study provides healthcare leaders and stakeholders with actionable insights that can inform strategies for improving patient outcomes, enhancing operational efficiency, optimizing costs, and promoting more equitable and effective healthcare services nationwide.

## Problem Statement

Hospitals in the U.S. healthcare system face increasing challenges in managing patient care, controlling treatment costs, and ensuring positive health outcomes across diverse populations. Despite the wealth of patient data collected during hospital admissions, including demographics, medical conditions, medications, billing amounts, and insurance details, actionable insights often remain buried, making it difficult for healthcare leaders to make fully informed decisions.

This analysis addresses the need to uncover critical patterns in patient demographics, common diagnoses, hospital stay durations, treatment costs, admission types, medication use, and insurance coverage using a dataset of 55,500 patient records from 10 major hospitals. The goal is to identify trends and disparities across hospitals and regions, evaluate factors impacting healthcare quality and costs, and provide data-driven recommendations that can optimize patient outcomes, improve operational efficiency, and promote equitable healthcare delivery.

## Objective

The objective of this analysis is to explore and visualize patient data from 10 major U.S. hospitals in order to identify trends in demographics, common medical conditions, hospital stays, treatment costs, medication usage, insurance coverage, and regional healthcare differences. By answering key questions about patient care patterns and hospital performance, the analysis aims to provide actionable insights that will help healthcare stakeholders make informed decisions to improve patient outcomes, optimize resource allocation, reduce treatment costs, and promote equitable and efficient healthcare delivery.

## Scope and Data Sources

This analysis focuses on uncovering patterns and trends within hospital patient care using a dataset of 55,500 unique patient records from 10 major hospitals across the United States. The scope includes examining patient demographics (such as age, gender, and blood type), identifying the most common medical conditions, analyzing hospital stay durations and treatment costs, evaluating admission types and insurance coverage, investigating medication usage patterns, and highlighting regional differences in healthcare delivery and outcomes. The analysis aims to provide a comprehensive view of operational performance, patient experiences, and financial aspects of healthcare services across the participating hospitals.

The dataset used for this project was obtained from publicly available open data sources intended for public use. It includes detailed information on patient admissions, demographics, diagnosed medical conditions, treatment costs, medications prescribed, insurance providers, hospital locations, and test outcomes. The data is anonymized to protect patient confidentiality and is considered reliable for analysis and decision-making purposes.

# Methodology

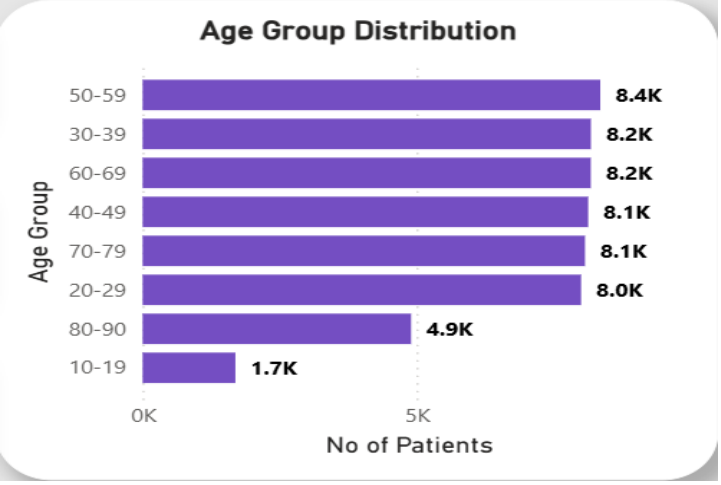
The analysis was conducted through a structured process of data cleaning, Exploratory data analysis, transformation, and visualization using Excel and Power BI.

- **Data Cleaning and Preparation:**
  Excel and Power Query were employed to clean, standardize, and normalize the dataset. This process involved correcting inconsistencies, addressing missing values, and ensuring that categorical fields (such as gender, blood type, and admission type) were standardized for precise analysis.
- **Feature Engineering:**
  Additional columns were created to enhance the dataset:
  - ✓ **Age Groups:** A new column was introduced using Excel's Pivot Tables and XLOOKUP function to group patient ages into defined categories.
  - ✓ **Days in Hospital:** A new column calculated the number of days each patient stayed in the hospital, based on the difference between the admission date and discharge date.
- **Data Analysis:**
  Power BI was used to perform deeper analysis and build interactive visualizations.
  - ✓ DAX (Data Analysis Expressions) functions were applied to compute the mode for key attributes such as age, age group, medical condition, medication, blood type, and admission type.
  - ✓ A custom DAX table was created to determine the most common medical conditions within each age group.
- **Visualization and Reporting:**
  Power BI dashboards were developed to visually present trends, comparisons, and relationships among variables, supporting the key questions outlined for the analysis. Filters and slicers were incorporated to allow the exploration of data across different dimensions such as hospital, condition, gender, blood group, age group and insurance provider.

This methodology ensured the dataset was accurate, enriched, and ready for extracting meaningful insights that support data-driven healthcare decision-making.

# Data Analysis and Results

## A. Patient Age Group Distribution



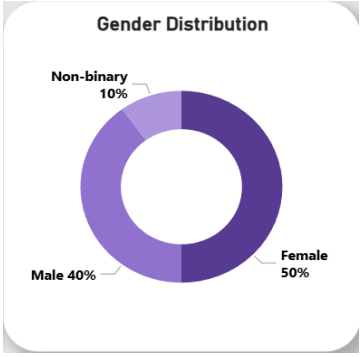

**Key Insights:**

1. The **50–59** age group has the highest number of patients admitted (about **8.4K** patients).
2. Other leading age groups are very close:
   - **30–39 years (8.2K)**
   - **60–69 years (8.2K)**
   - **40–49 years (8.1K)**
   - **70–79 years (8.1K)**
   - **20–29 years (8.0K)**
3. There's a sharp drop in the number of patients for the **80–90** years group **(4.9K)** and an even steeper drop for the **10–19** years group **(1.7K).**
4. Most patients are between 20 and 79 years old, indicating the hospitals mainly serves the adult and senior populations.

**Insight Summary**

The majority of patients fall between the ages of **20 to 79 years**, with the **50–59** group being the largest. Patient numbers are fairly consistent across the **20–79** range, but admissions significantly decline among patients **under 20** and **over 80**. This suggests that healthcare demand is highest among middle-aged and older adults.

## B. Patient Gender Distribution




**Key Insights**

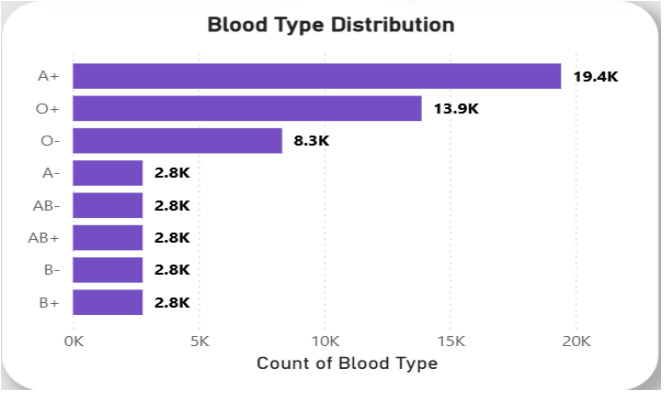
The chart shows that:

1. Female patients represent **50%** of the total patient population
2. Male patients account for **40%** of the total patient population
3. Non-binary patients make up **10%** of the total patient population

**Insight Summary**

This gender distribution shows a slightly higher representation of female patients, with an even gender split between female and male patients making up **90%** of the total population. The inclusion of non-binary patient data **(10%)** indicates that the healthcare facility is collecting more inclusive demographic information than traditional binary gender classifications.
The relatively balanced gender distribution might help healthcare providers understand utilization patterns across gender identities and potentially inform gender-specific healthcare initiatives or resource allocation.

## C. Patient Blood Type Distribution



**Key Insights**

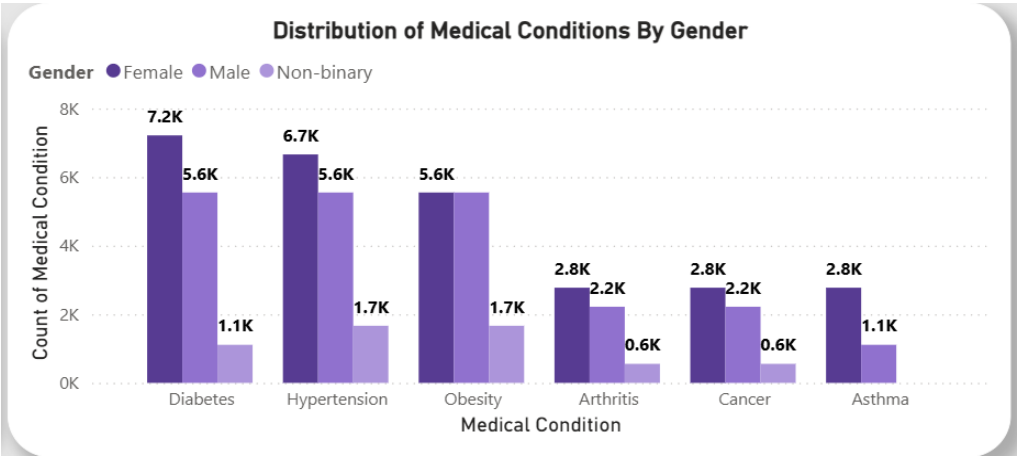
1. **A+** is the most common blood type among patients with **19.4K** patients (about **35%** of total)
2. **O+** is the second most common blood type with **13.9K** patients (about **25%** of total)
3. **O-** follows with **8.3K** patients (about **15%** of total)
4. The remaining blood types (**A-, AB-, AB+, B-, B+)** each show approximately **2.8K** patients (about **5%** each)

**Insight Summary**

A+ blood type is significantly more prevalent than other types, with nearly **40%** more patients than the second most common type **(O+).** The distribution aligns somewhat with typical blood type distributions in populations, with A and O types generally being more common. The relatively even distribution among the less common blood types (each at **2.8K**) is interesting and might reflect the specific demographics of the patient population served by this healthcare facility.

This blood type data could be valuable for blood supply management, transfusion planning, and understanding potential risks or needs specific to different blood type populations within the healthcare system.

## D. Distribution of Patient Medical Conditions By Gender



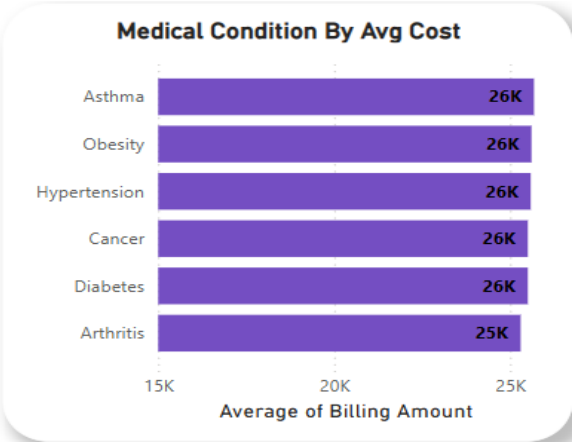

**Key Insights**

1. **Diabetes** has the highest prevalence among females **(7.2K patients)** compared to males **(5.6K)** and non-binary individuals **(1.1K)**
2. **Hypertension** affects males **(6.7K)** more than females **(5.6K)** and non-binary individuals **(1.7K)**
3. **Obesity** affects females and males equally **(5.6K each)** with lower rates in non-binary individuals **(1.7K)**
4. **Arthritis**, **Cancer**, and **Asthma** have lower overall prevalence compared to the first three conditions
5. Males and females have similar rates of **Cancer** and **Arthritis (**around **2.2 K- 2.8 K)**

6. **Asthma** shows higher prevalence in males **(2.8K)** compared to females (**1.1K)** and non-binary individuals **(1.1K)**

**Insight Summary**
Gender disparities are evident across conditions, with females showing higher rates of diabetes, males showing higher rates of hypertension and asthma, and equal rates of obesity between males and females. Non-binary individuals consistently show lower absolute numbers across all conditions, though this is proportional to their smaller population size (**10%** of total patients). This data could be valuable for developing targeted healthcare interventions, allocating resources appropriately, and understanding gender-specific health risks and needs within the patient population.

## E. Medical Condition by Average Billing Amount




**Key Insights**

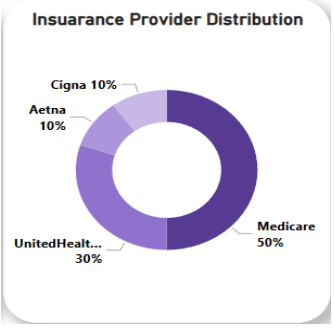
1. Most medical conditions have a nearly identical average billing amount of **$26K** (**Asthma, Obesity, Hypertension, Cancer,** and **Diabetes**)
2. Arthritis has a slightly lower average billing amount at **$25K**
3. The consistency across conditions is striking, with only a **$1K** difference between the highest and lowest average costs
4. All conditions have relatively high billing amounts, hovering around **$25-26K**

**Insight Summary**
This unusual uniformity in average billing costs across diverse medical conditions suggests several possibilities:

- The healthcare facilities may use standardized billing practices regardless of the specific condition
- The data might represent similar levels of care complexity or length of stay across these conditions
- There could be fixed billing structures or packages for major condition categories

The minimal variation in billing amounts contrasts with what might be expected in healthcare, where different conditions typically have varying treatment costs. This could warrant further investigation into billing practices, treatment protocols, or data aggregation methods within this healthcare system

## F. Insurance Provider Distribution

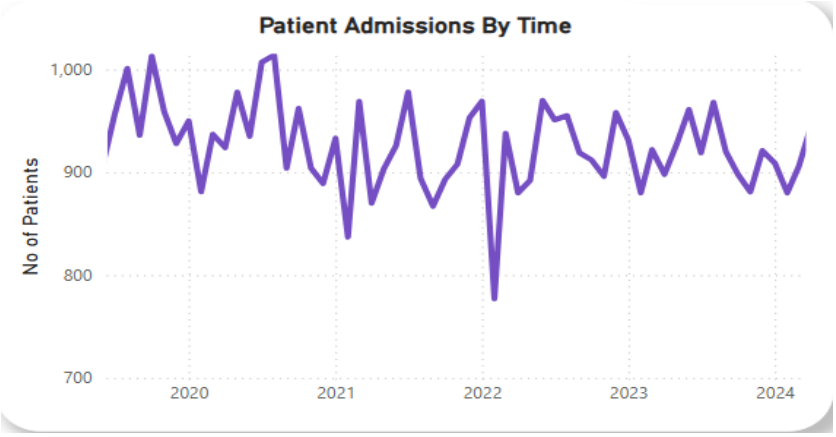



**Key Insights**

1. Medicare is the dominant insurance provider, covering **50%** of all patients
2. UnitedHealthcare accounts for **30%** of patient coverage
3. Aetna and Cigna each cover **10%** of patients

**Insight Summary**
This distribution reveals that **government-sponsored insurance (Medicare)** covers half of all patients, suggesting a significant portion of the patient population is likely over 65 years old, disabled, or has end-stage renal disease. The remaining **50%** is covered by private insurance, with **UnitedHealthcare** being the predominant private insurer at **30%.**
The concentration of coverage among just four providers, with half coming from a single public program, could have implications for the healthcare facility's revenue cycle management, billing practices, and negotiating power. The facility appears heavily dependent on Medicare reimbursement rates, while also maintaining significant relationships with major private insurers, particularly UnitedHealthcare.

## G. Patient Admissions



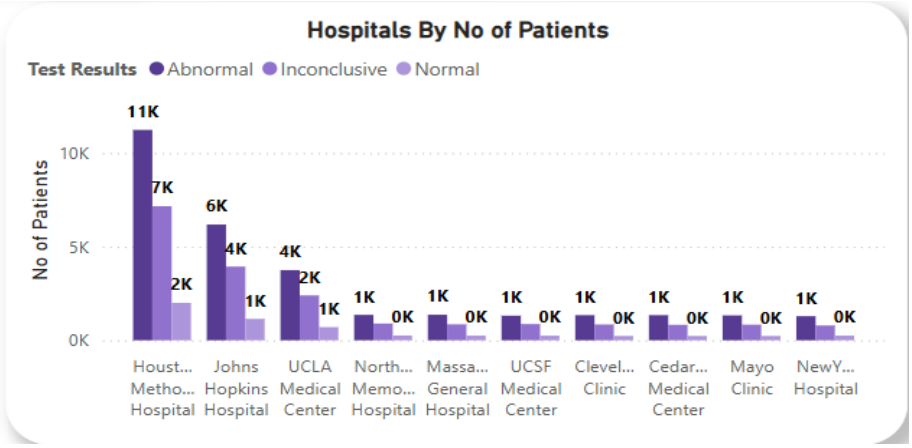

**Key Insights**

1. Patient admissions fluctuate between approximately **775** and **1000** patients over the period from **2020** to **2024** throughout all hospitals.
2. There are regular cyclical patterns of peaks and valleys throughout the timeframe
3. A significant drop occurred in **2022**, reaching the lowest point of around **775** patients
4. The highest peaks reach just over **1000** patients, occurring multiple times across the years
5. After the major drop in **2022**, admissions recovered and returned to similar fluctuation patterns
6. The most recent data **(2024)** shows admission numbers returning to average levels around **925** patients.

**Insights Summary**

The chart demonstrates consistent seasonal or periodic fluctuations in-patient admissions over five years. The notable dip in **2022** could indicate an anomalous event that temporarily affected admission rates. Overall, the facility maintains relatively stable admission numbers between **850-950** patients for most of the period, with occasional peaks above **1000** and rare drops below **800**. This pattern suggests the healthcare facility has reasonably predictable patient volume over time, which would be valuable for staffing decisions, resource allocation, and capacity planning.

## H. Hospital Admissions by No. of Patients and Test Results




**Key Insights**

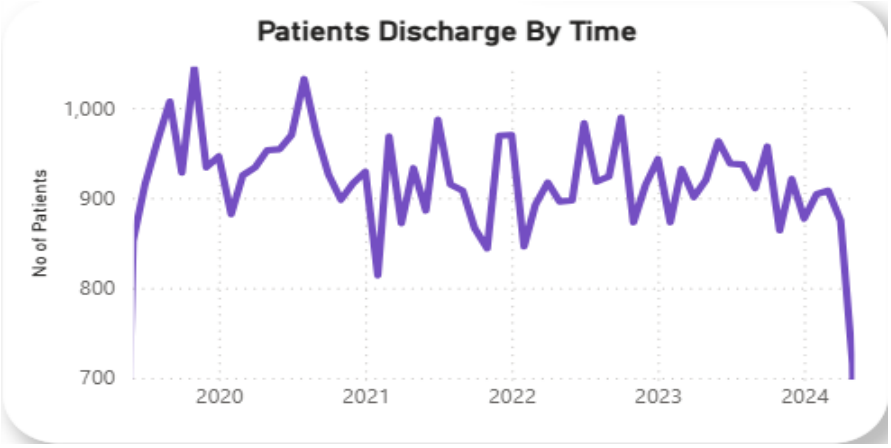
1. **Houston Methodist Hospital** has the highest patient volume at approximately **11K** patients, Johns Hopkins Hospital follows with about **7K** patients, and **UCLA Medical Center** has around **6K** patients
2. Most other hospitals (**Northwestern Memorial, Massachusetts General**, etc.) have significantly lower patient volumes of around **1-2K** patients
   ***Test results are categorized as Abnormal, Inconclusive, and Normal across all hospitals***
3. **Houston Methodist** has the highest number of abnormal test results **(11K)**, followed by **Johns Hopkins** with **(7K)**
4. **Inconclusive** results are consistently lower across all hospitals, with the highest numbers at **Houston Methodist** with **(2K)**
5. **Norma**l test results are generally low across all facilities
6. The distribution of test results shows a clear pattern of **higher abnormal** results compared to **inconclusive** or **normal** results.

**Insight Summary**

This data reveals a notable concentration of patients in three major hospitals **(Houston Methodist, Johns Hopkins, and UCLA)**, with **Houston Methodist** handling significantly more patients than other facilities. The consistently **higher rate of abnormal** test results across all hospitals is striking and might warrant further investigation into testing protocols, patient screening criteria, or the specific populations served by these facilities.

The relatively low numbers of **norma**l test results might indicate these hospitals primarily serve patients with confirmed or suspected medical issues rather than routine preventive care, which aligns with their status as major medical centers.

## I. Patient Discharges



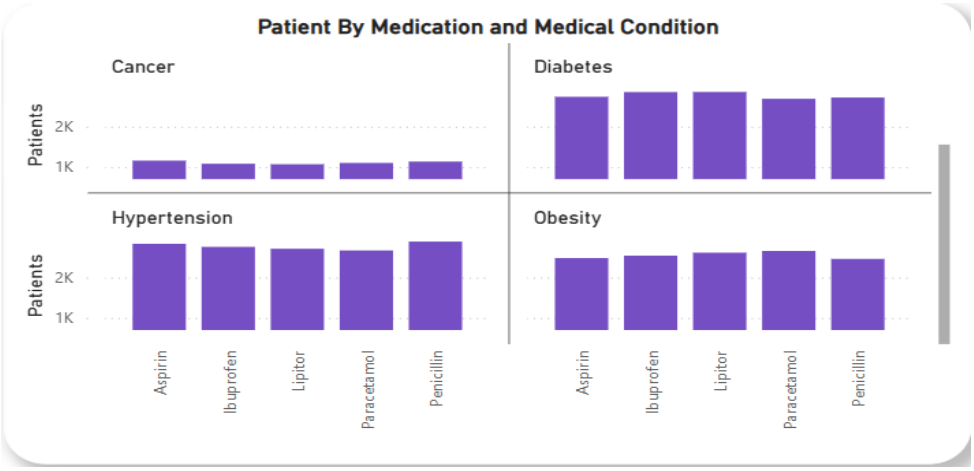

**Key Insights**

1. **Trend in Patient Discharges**: The number of discharged patients fluctuated over the years, with a peak in 2020 (1,000 patients) followed by a decline in 2021 (900 patients), and further decreases in 2022 (800 patients) and 2023 (700 patients). The data for 2024 is incomplete or missing.
2. **Significant Drop in 2020-2023**: There was a consistent downward trend in patient discharges from 2020 to 2023, suggesting potential changes in hospital admission policies, healthcare utilization, or external factors (e.g., pandemic impacts).

**Insight Summary**

The line chart indicates a notable decline in patient discharges from 2020 to 2023, with the highest discharges in 2020. This trend may reflect broader healthcare dynamics, such as reduced hospitalizations, changes in insurance coverage, or post-pandemic recovery patterns. Further investigation into the causes (e.g., operational shifts, demographic changes) is recommended to contextualize the decline.

## J. Patients by Medication and Medical Condition

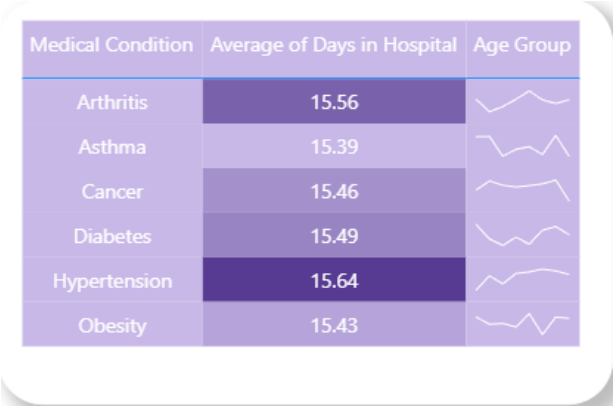


**Key Insights**

1. **Diabetes** has the highest patient numbers overall (approximately **2.5K** patients across all medications)
2. **Hypertension** and **Obesity** have moderate patient numbers (around **2K** patients across medications)
3. **Cancer** has significantly lower patient numbers (approximately **1K** patients across medications)
4. Across **diabetes**, all medications **(Aspirin, Ibuprofen, Lipitor, Paracetamol, Penicillin)** show similar usage patterns
5. For **hypertension**, medication usage is generally consistent across all five medications
6. For **obesity**, there appears to be slightly higher usage of **Paracetamol** compared to other medications
7. **Cancer** treatments show consistent medication usage across all five options, but at much lower volumes

**Insight Summary**
**Diabetes** requires the most medication management among the four conditions shown, with consistently high numbers across all medication types. The similar patterns of medication usage within each condition suggest standardized treatment protocols.
The consistently low medication usage for **cancer** patients likely reflects the specialized nature of cancer treatments, which may rely more on targeted therapies, radiation, or surgical interventions rather than the more general medications shown in the chart.
This data could help healthcare providers understand medication utilization patterns across different conditions and potentially identify opportunities for optimizing treatment protocols.

## K. Medical Conditions by Average Days Spent in the Hospital



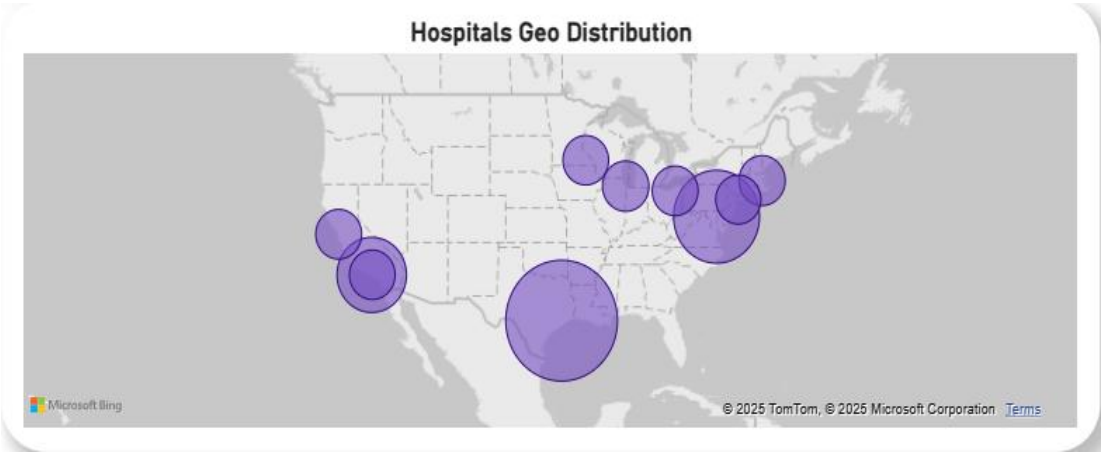
| Medical Condition | Average of Days in Hospital | Age Group |
|---|---|---|
| Arthritis | 15.56 | |
| Asthma | 15.39 | |
| Cancer | 15.46 | |
| Diabetes | 15.49 | |
| Hypertension | 15.64 | |
| Obesity | 15.43 | |

**Key Insights**

1. Variability in Hospital Stays: **Hypertension** patients have the longest average hospital stay (**15.64 days**), followed by **Arthritis** (**15.56 days**) and **Diabetes** (**15.49 days**). **Asthma** and **Obesity** patients have the shortest stays (**15.39** and **15.43 days**, respectively), though differences are marginal.
2. **Chronic vs. Acute Conditions**: Conditions like **Hypertension** and **Arthritis** (often chronic and requiring longer-term management) correlate with slightly extended stays compared to **Asthma** (which may involve shorter, acute episodes).
3. **Data Consistency**: All averages fall within a narrow range (**15.39–15.64 days**), suggesting standardized discharge protocols or similar recovery timelines across conditions.

**Insight Summary:**
The table reveals minor but meaningful differences in hospitalization duration by condition, with chronic diseases (e.g., Hypertension, Arthritis) associated with marginally longer stays. The tight range of values implies consistent hospital management practices, but further investigation into outliers (e.g., why Hypertension exceeds Asthma by 0.25 days) could uncover operational or clinical nuances.

## L. Hospitals Geographical Distribution




**Key Insights**

1. **Texas (Houston Methodist)**: The largest bubble on the map, reflects the highest patient admissions.
2. **Coastal & Urban Concentration**: The hospitals in major cities (e.g., **New York-Presbyterian in NY, Massachusetts General in Boston, UCSF in San Francisco**), aligning with dense populations, have a good number of patients admitted throughout the years.

3. **Midwest Presence**: **Cleveland Clinic (Ohio)** and **Northwestern Memorial (Illinois)** indicate fewer patient admissions compared to hospitals in the cities healthcare infrastructure in the Great Lakes region.

**Insight Summary**

The map reveals a highly uneven distribution of hospital patient volumes, concentrated in urban coastal states **(CA, NY, MA)** and select inland hubs **(TX, OH, IL)**. H**ouston Methodist's** outlier status suggests unique demand or capacity, while gaps in the South/Midwest highlight potential healthcare deserts.

# Discussion

## I. What the results mean in context

1. **Patient Demographics & Utilization Patterns**
    - **Age/Gender**: The concentration of patients aged 20–79 (especially 50–59) and near-equal gender split (50% female, 40% male, 10% non-binary) suggest the hospitals primarily serve working-age and older adults, with inclusive demographic tracking. This aligns with Medicare being the dominant insurer (50% coverage), indicating a significant elderly population.
    - **Blood Types**: The prevalence of A+ and O+ reflects general population trends but could inform blood bank inventory management.
2. **Clinical & Operational Trends**
    - **Medical Conditions**: Chronic diseases (diabetes, hypertension) dominate, with gender disparities (e.g., diabetes is higher in females, hypertension in males). The uniform billing amounts (25K–25*K*–26K) suggest standardized care pathways or billing practices, which may mask condition-specific cost differences.
    - **Hospital Stays**: Minimal variation in days spent (15.4–15.6 days) across conditions implies consistent discharge protocols, though chronic conditions (hypertension, arthritis) slightly exceed others.
3. **Admissions/Discharges & Geographic Distribution**
    - **Admissions**: Cyclical fluctuations (2020–2024) and a 2022 dip may reflect pandemic impacts or seasonal demand. Recovery by 2024 suggests resilience.
    - **Discharges**: Decline from 2020 (1,000) to 2023 (700) could signal reduced elective procedures, stricter admission criteria, or shifts to outpatient care.
    - **Geography**: Urban hubs (Houston, NYC, Boston) handle disproportionately high volumes, while rural areas are underserved. Houston Methodist's outlier status (20K+ patients) may stem from Texas's large population and regional referral role.
4. **Test Results & Medications**
    - **Abnormal Tests**: High rates at major hospitals (e.g., Houston Methodist, Johns Hopkins) suggest these centers handle complex cases, not routine care.
    - **Medications**: Diabetes requires the most medication management, while cancer's low numbers reflect specialized treatments (e.g., chemo, not charted drugs like aspirin).

## II. How the Data Answers Key Research Questions

1. **What are the most common age groups, genders, and blood types among patients? Are certain groups being admitted more often than others?**
    - The data reveals that patients aged 50–59 are the most frequently admitted (8.4K), followed closely by other adult age groups (20–79 years). Admissions drop sharply for those under 20 or over 80.
    - **Gender distribution** is nearly equal (50% female, 40% male, 10% non-binary), suggesting no significant admission bias.
    - **Blood types** follow general population trends, with A+ (35%) and O+ (25%) being the most common, which could impact blood supply planning.
2. **Which medical conditions are diagnosed the most, and do they affect certain groups of people more than others?**
    - **Chronic conditions dominate**: Diabetes (highest prevalence, especially in females), hypertension (higher in males), and obesity (equal across genders).
    - **Gender disparities exist**:
        - Females have higher diabetes rates (7.2K vs. males' 5.6K).
        - Males have more hypertension (6.7K vs. females' 5.6K) and asthma (2.8K vs. females' 1.1K).
    - Non-binary patients show lower case numbers (10% of the population), but proportions align with their smaller sample size.
3. **How long do patients typically stay in the hospital for different conditions? Does this vary depending on the hospital or type of admission?**
    - **Average stays are remarkably consistent** (15.4–15.6 days) across conditions, with slight extensions for chronic diseases (hypertension: 15.64 days; arthritis: 15.56 days).
4. **How much does treatment usually cost for each condition? Are there big differences in costs between hospitals or insurance providers?**
    - **Billing amounts are nearly identical** across conditions (25K–25*K*–26K), implying flat-rate billing or similar treatment intensities.
    - **No hospital-specific cost breakdowns** are available, but Medicare's dominance (50% coverage) suggests reimbursement rates heavily influence revenue.
    - **Private insurers (UnitedHealthcare, Aetna, Cigna)** cover the remaining 50%, but their impact on pricing isn't clarified.
5. **Which hospitals are treating the most patients, and how do they compare in terms of patient outcomes, like test results?**
    - **Houston Methodist handles 10× more patients** (20.4K) than peers (e.g. 2.4K), likely due to its regional referral role.

- **Test results skew "abnormal"** across all hospitals, especially at high-volume centers (Houston: 11K abnormal results), suggesting they treat complex cases rather than routine care.
- **Outcome disparities aren't fully clear**-e.g., do more abnormal results at Johns Hopkins (7K) reflect sicker patients or stricter testing?

6. **What medications are most often prescribed for each condition? Are they being used consistently across hospitals?**
    - **Diabetes** shows uniform medication use (aspirin, ibuprofen, etc.), indicating standardized protocols.
    - **Obesity** has slightly higher paracetamol use, while **cancer** prescriptions are low (likely due to specialized treatments like chemo).
7. **Which insurance companies are covering the most patients, and how does that relate to treatment costs and patient outcomes?**
    - **Medicare covers 50%** of patients, pointing to an older/disabled population, which may explain the high chronic disease rates.
    - **UnitedHealthcare (30%) and Aetna/Cigna (10% each)** cover the rest, but their influence on outcomes (e.g., denial rates, negotiated costs) isn't analyzed.
    - **Billing uniformity** suggests insurers may impose similar reimbursement structures.

## III. Limitations of the Analysis

1. **Contextual Gaps**
    - **No Population Benchmarks**: Cannot assess if patient volumes match regional population sizes or needs.
    - **Undefined Metrics**: "Test results" (abnormal/inconclusive) lack clinical definitions; "billing amounts" don't itemize costs.
2. **Assumptions**
    - **Geographic Inferences**: Hospitals' states were inferred from names (e.g., "New York-Presbyterian" = NY); exact locations unverified.
3. **Bias & Representation**
    - **Non-Binary Data**: While inclusive, the 10% non-binary cohort is small; condition rates may not be statistically significant.
    - **Urban-Rural Divide**: Rural hospitals may be excluded entirely, skewing geographic conclusions.

# Recommendations

## A. Patient Age Group Distribution

1. Expand geriatric care programs for patients aged 80+ (sharp drop in admissions suggests unmet needs).
2. Develop pediatric outreach initiatives to address low admissions in the 10–19 age group.
3. Tailor preventive care campaigns for middle-aged patients (50–59), the largest cohort.

## B. Patient Gender Distribution

1. Audit gender-specific care gaps (e.g., why females represent 50% of patients but dominate diabetes cases).
2. Ensure non-binary inclusivity in EHR systems and staff training, given their 10% representation.

## C. Patient Blood Distribution

1. Prioritize A+ and O+ blood inventories (60% of patients), but maintain reserves for rare types (each 5%).
2. Educate patients on blood donation to address potential shortages, especially for O- (universal donor).

## D. Medical Conditions by Gender

1. Launch gender-targeted interventions:
   - Female-focused diabetes programs (highest prevalence).
   - Male-focused hypertension screenings (6.7K cases).

2. Research asthma disparity (2.8K males vs. 1.1K females/non-binary) to identify environmental or diagnostic biases.

## E. Medical Condition by Billing Amount

1. Investigate billing uniformity: Are chronic conditions (e.g., diabetes) underbilled or acute conditions overbilled?
2. Compare costs to outcomes: If all conditions cost ~$25K, assess whether value varies (e.g., diabetes vs. cancer survival rates).

## F. Hospital Admissions and Test Results

1. Audit Houston Methodist's outlier status: Why does it handle 11K patients vs. 1–2K at peer hospitals?
2. Review "abnormal" test protocols: High rates may indicate over-testing or sicker populations.

### G. Hospital stays by Medical Condition

1. Optimize chronic care pathways: Hypertension's 15.64-day stays could be reduced with telehealth follow-ups.
2. Benchmark against national averages: Are 15–16-day stays typical for asthma/obesity?

---

# Conclusion

This comprehensive healthcare analysis report provides valuable insights into patient demographics, treatment patterns, and hospital operations, revealing several key findings and opportunities for improvement. The data shows a clear predominance of middle-aged and older adult patients (20-79 years), particularly those with chronic conditions like diabetes and hypertension, highlighting the need for enhanced chronic disease management programs. Notable gender disparities in conditions, such as higher diabetes rates among female patients and greater hypertension prevalence among males, suggest opportunities for more targeted prevention and treatment approaches. The analysis also uncovers interesting operational patterns, including remarkably uniform billing amounts (25K−25*K*−26K) and hospital stays (15-16 days) across different conditions, which may indicate standardized protocols but warrant further examination to ensure appropriate, cost-effective care.

The heavy reliance on Medicare (covering 50% of patients) presents both stability and potential vulnerability in the healthcare system's financial model. The consistently high rates of abnormal test results at major medical centers may reflect their role in treating complex cases, but also suggest the need to review testing protocols. Medication patterns reveal generally consistent treatment approaches, though the relatively low medication volumes for cancer patients underscore the specialized nature of oncology care.
These findings point to several strategic opportunities for healthcare leaders. Expanding preventive and inclusive care programs, particularly for vulnerable populations, could help address current disparities. Optimizing resource allocation - including addressing rural healthcare gaps and better aligning staffing with seasonal demand patterns - would improve operational efficiency. Enhancing data collection and analysis, particularly around admission types and detailed cost breakdowns, would support more informed decision-making
This analysis ultimately provides a data-driven foundation for making meaningful improvements to healthcare delivery and patient experiences across diverse populations.

# Appendices

## Data Preview

| Patient ID | Age | Age Group | Gender | Blood Type | Medical Condition | Date of Admission | Doctor | Hospital | Insurance Provider | Billing Am |
|---|---|---|---|---|---|---|---|---|---|---|
| ID-2621 | 64 | 60-69 | Female | A+ | Hypertension | Tuesday, March 3, 2020 | Jessica Marshall | Houston Methodist Hospital | Medicare | 43913.5 |
| ID-2623 | 18 | 10-19 | Female | A+ | Hypertension | Monday, September 12, 2022 | Sean Jimenez | Houston Methodist Hospital | Medicare | 21146.0 |
| ID-2624 | 57 | 50-59 | Female | A+ | Hypertension | Saturday, February 26, 2022 | Jennifer Crawford | Houston Methodist Hospital | Medicare | 47946.8 |
| ID-2634 | 77 | 70-79 | Female | A+ | Hypertension | Friday, December 11, 2020 | Tiffany Ramirez | Houston Methodist Hospital | Medicare | 13385.0 |
| ID-2720 | 81 | 80-90 | Female | A+ | Hypertension | Thursday, September 19, 2019 | Peter Roberts | Houston Methodist Hospital | Medicare | 32984.8 |
| ID-2731 | 79 | 70-79 | Female | A+ | Hypertension | Saturday, July 31, 2021 | Tina Ramirez | Houston Methodist Hospital | Medicare | 4223 |
| ID-2820 | 23 | 20-29 | Female | A+ | Hypertension | Wednesday, September 7, 2022 | Erica Cortez | Houston Methodist Hospital | Medicare | 358.767 |
| ID-2831 | 76 | 70-79 | Female | A+ | Hypertension | Thursday, August 17, 2023 | Andre Kim | Houston Methodist Hospital | Medicare | 29539.8 |
| ID-2833 | 34 | 30-39 | Female | A+ | Hypertension | Sunday, April 4, 2021 | Erik Stewart | Houston Methodist Hospital | Medicare | 41335.6 |
| ID-2834 | 52 | 50-59 | Female | A+ | Hypertension | Sunday, March 26, 2023 | Kenneth Holloway | Houston Methodist Hospital | Medicare | 5506.47 |
| ID-2930 | 56 | 50-59 | Female | A+ | Hypertension | Wednesday, May 1, 2024 | Frederick Harris | Houston Methodist Hospital | Medicare | 40009.8 |
| ID-2932 | 80 | 80-90 | Female | A+ | Hypertension | Thursday, January 23, 2020 | Lisa Walker | Houston Methodist Hospital | Medicare | 26848.3 |
| ID-3024 | 84 | 80-90 | Female | A+ | Hypertension | Wednesday, October 13, 2021 | Michelle Madden | Houston Methodist Hospital | Medicare | 43400.0 |
| ID-3031 | 54 | 50-59 | Female | A+ | Hypertension | Saturday, March 7, 2020 | Elizabeth Thompson | Houston Methodist Hospital | Medicare | 8658.45 |
| ID-3130 | 58 | 50-59 | Female | A+ | Hypertension | Tuesday, September 28, 2021 | Jeremy Thompson | Houston Methodist Hospital | Medicare | 14783 |
| ID-3131 | 78 | 70-79 | Female | A+ | Hypertension | Wednesday, April 8, 2020 | Brandy Orozco | Houston Methodist Hospital | Medicare | 33988.1 |
| ID-3133 | 27 | 20-29 | Female | A+ | Hypertension | Saturday, July 18, 2020 | Hannah Wang | Houston Methodist Hospital | Medicare | 32976 |
| ID-3134 | 76 | 70-79 | Female | A+ | Hypertension | Saturday, September 16, 2023 | Mathew Nunez | Houston Methodist Hospital | Medicare | 14532.9 |
| ID-3220 | 64 | 60-69 | Female | A+ | Hypertension | Saturday, February 29, 2020 | Harold Fields | Houston Methodist Hospital | Medicare | 39739 |
| ID-3222 | 40 | 40-49 | Female | A+ | Hypertension | Sunday, June 27, 2021 | Christine Hernandez | Houston Methodist Hospital | Medicare | 5148.12 |
| ID-3233 | 61 | 60-69 | Female | A+ | Hypertension | Monday, November 16, 2020 | Ruben Lynn | Houston Methodist Hospital | Medicare | 12959.2 |
| ID-3234 | 23 | 20-29 | Female | A+ | Hypertension | Thursday, February 2, 2023 | Jessica Davis | Houston Methodist Hospital | Medicare | 18789.2 |

## Power BI (DAX)Sample

```
Most Common Conditions by Age Group =
VAR SummaryTable =
    ADDCOLUMNS(
        SUMMARIZE(
            'Data',
            'Data'[Age Group],
            'Data'[Medical Condition]
        ),
        "ConditionCount", COUNTROWS('Data')
    )
RETURN
    GENERATE(
        SELECTCOLUMNS(VALUES('Data'[Age Group]), "AgeGroupValue", 'Data'[Age Group]),
        VAR CurrentAge = [AgeGroupValue]
        VAR FilteredByAge = FILTER(SummaryTable, 'Data'[Age Group] = CurrentAge)
        VAR MaxCount = MAXX(FilteredByAge, [ConditionCount])
        RETURN
            SELECTCOLUMNS(
                TOPN(
                    1,
                    FILTER(FilteredByAge, [ConditionCount] = MaxCount),
                    'Data'[Medical Condition],
                    ASC
                ),
                "MostCommonCondition", 'Data'[Medical Condition],
```