**Project Report**

**Data Visualization with Tableau**

**INFS - 420**

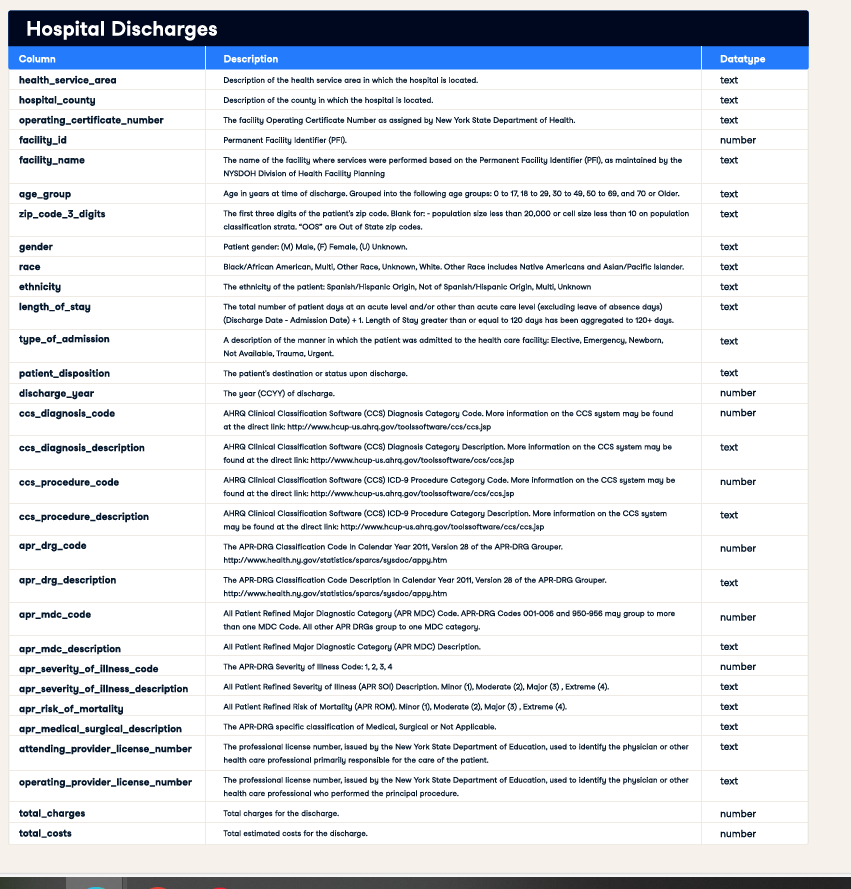
**Hospital Performance & Patient Insights Analysis**

**A Data-Driven Approach using Tableau**

**Submitted by: Mariam Iftikhar**

**Dataset Description**

The dataset has been taken from Datacamp that includes anonymous ***New York State-wide Hospital discharge data*** for 2016-2023. The main reason for the Hospital stays of dataset patients was Elective hip replacement surgery. The dataset is a single table with 30 columns. Each row in the dataset represents a single inpatient stay, from their admission to discharge date. This includes a unique identifier for facility, a grouping of patient age, patient's disposition, the diagnosis description, severity of illness, and risk of mortality along with length of stay, measured in total days, and total costs, attributed to each hospital stay.



**Business Scenario**

As a Healthcare Data Analyst working for the consulting company called HealthStat for a hospital network or health department, the task is to analyze hospital utilization, patient demographics, disease patterns, cost efficiency, and healthcare quality to improve operational efficiency, patient care, and financial performance.

An important indicator of the efficiency of hospital management is the length of stay (or LOS). It is calculated as the total duration of days for a patient stay in the hospital. A shorter LOS is often desirable in hospital operations. Primarily, shorter LOS means costs can be lowered. Reducing LOS can also release capacity in the system and improve throughput, enabling hospitals to serve more patients. Many factors can impact LOS, Patient age, health status, the type of procedure (or surgery), whether there were any complications, and the size of the hospital are some common factors.

**KEY BUSINESS QUESTIONS TO ANSWER:**

**Hospital Utilization & Patient Demographics:**

* Which age groups experience the highest rates of hospital admissions?
* How do hospital admissions differ by gender, race, and ethnicity?
* Which counties or ZIP codes show the highest levels of hospital utilization?
* How have trends in admissions evolved over time for various diagnoses?
* What are the most common diagnoses and procedures?
* How does length of stay vary across different patient demographics?

**Disease Patterns & Severity of Illness:**

* What are the most frequent diagnoses that lead to hospital admissions?
* How does the length of stay differ based on diagnosis and illness severity?
* What is the distribution of mortality risk scores among patients who are admitted?
* Which procedures or conditions are associated with the longest hospital stays?
* Which diseases or conditions have the highest risk of mortality?

**Cost & Revenue Optimization:**

* How do hospital charges compare to actual costs across different facilities?
* Are there any high-cost hospitals that need cost-cutting measures?
* How do total charges stack up against total costs for different procedures?
* Which diagnoses and procedures result in the highest expenses?
* In what ways does the length of stay affect hospital charges when it comes to diagnosis?

**Analysis and Insights**

**KEY INSIGHTS:**

**Hospital Utilization & Patient Demographics:**

Individuals aged 50 and older account for most hospital visits, with a slight predominance of females over males. The 30 to 49 age group experiences significantly fewer hospitalizations than their older counterparts. Meanwhile, the pediatric (0 to 17) and young adult (18 to 29) groups show the lowest hospitalization rates, indicating they are admitted less frequently.

Among the admissions, White patients represent the largest group (21,853), followed by those identified as Other Race (2,249) and Black/African American (2,227). In terms of ethnicity, most patients are Not Spanish/Hispanic (24,764), while Spanish/Hispanic patients make up a smaller portion with 1,392 admissions.

Manhattan leads with the highest number of admissions at 8,250, followed by Erie with 2,426 and Nassau with 2,006.

In 2022, there was a notable increase in patient counts and total costs, suggesting a possible rise in hospital usage. Projected figures for 2023 indicate a decrease in patient counts compared to 2022, yet total costs remain elevated, implying higher costs per patient. The prediction intervals for 2023 reveal uncertainty regarding both patient counts and total costs.

Osteoarthritis stands out as the most prevalent diagnosis, with 23,758 cases, followed by complications related to devices, implants, or grafts, which account for 1,568 cases. Hip replacement procedures, whether total or partial, are the most common, totaling 26,286 cases, while other joint procedures are significantly less frequent.

**A screenshot of a graph

AI-generated content may be incorrect.**Younger patients (ages 0-17) and older patients (70+) tend to have longer average hospital stays, at 5.69 and 2.99 days respectively, whereas middle-aged patients have shorter stays. Female patients have a slightly longer average stay of 2.79 days compared to male patients, who average 2.54 days. Among racial groups, Black/African American patients have the longest average stays at 3.13 days, while White patients have the shortest at 2.58 days.

**Disease Patterns & Severity of Illness:**

Osteoarthritis is the leading cause of hospital admissions, with 23,758 cases reported. This is followed by complications related to devices, implants, or grafts, which account for 1,568 cases, and other bone diseases and musculoskeletal deformities, totaling 700 cases.

The data indicates that patients experiencing complications from devices, implants, or grafts tend to have longer hospital stays, averaging 3.84 days, compared to those with osteoarthritis, who average 2.53 days. When examining severity, patients classified with "Extreme" severity of illness have notably longer stays, averaging 9.83 days, in contrast to those with "Minor" severity, who average 2.46 days.

A significant number of patients, 22,621, are categorized as having a "Minor" risk of mortality, while only 100 patients fall into the "Extreme" risk category. This distribution aligns with the elective nature of many orthopedic procedures included in the dataset.

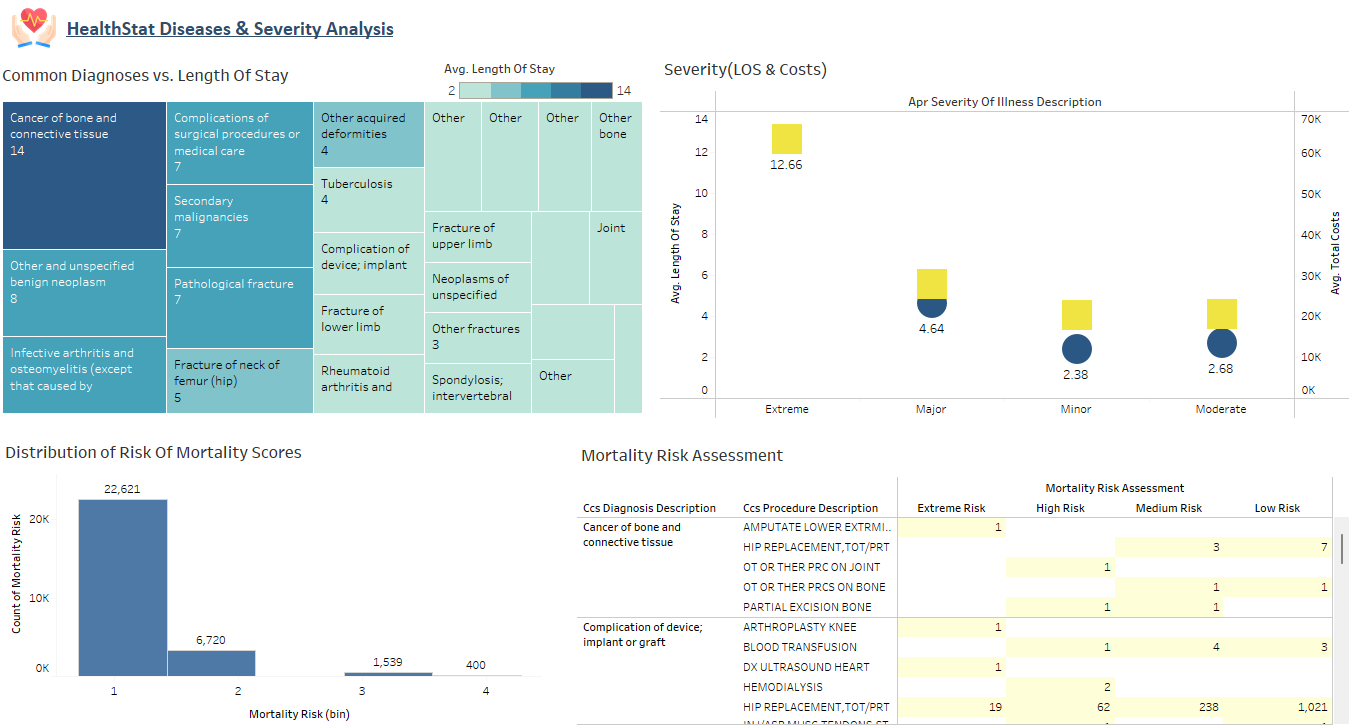
Blood transfusions are linked to the longest average hospital stays, at 4.12 days, followed by other therapeutic procedures on joints, which average 3.60 days. Hip replacements, the most common procedure, have an average stay of 2.65 days.

When considering actual mortality outcomes based on patient disposition labeled as "Expired":

- Fractures of the lower limb have the highest mortality rate at 25%

- Pathological fractures follow at 1.82%

- Fractures of the neck of the femur or hip have a mortality rate of 1.72%

These findings indicate that while orthopedic procedures like hip replacements are prevalent, fractures, particularly those of the lower limb and hip, present significantly higher mortality risks, likely due to complications and comorbidities associated with these patients.

**Cost & Revenue Optimization:**

There is a notable range of cost-to-charge ratios across different facilities, varying from around 0.20 to more than 2.25. This suggests that hospitals have significant differences in how much they charge compared to their actual costs. Hospitals with high-cost ratios (above 0.7) may warrant further investigation to explore potential cost-reduction strategies, as their charges represent a large portion of their costs.

The most common diagnoses by average total charges, such as “Cancer of bone and connective tissue,” indicate areas with substantial billing, while the leading procedures also reflect high-cost care. This information can help pinpoint which conditions and treatments are likely to contribute most to overall healthcare spending.

Certain procedures, including "DX ULTRASOUND HEART," "AMPUTATE LOWER EXTREMITY," and "ARTHROPLASTY KNEE," show very high average charges. However, their average total costs vary, suggesting that different procedures have distinct cost structures or profit margins. Among the diagnoses, conditions like "Cancer of bone and connective tissue" and "Complication of device; implant or graft" rank highest in charges, indicating that these cases could significantly increase hospital expenses.

There is a positive correlation suggesting that longer hospital stays are somewhat linked to higher total charges.

**A screenshot of a computer

AI-generated content may be incorrect.**

**Recommendations**

* **Benchmarking and Process Optimization:**

Hospitals with high cost-to-charge ratios, particularly those highlighted in the high-cost analysis, should take a closer look at their internal cost structures and operational efficiency.

* **Focus on Length of Stay Management:**

Given the moderate link between length of stay and charges, hospitals might explore strategies to enhance care pathways and minimize unnecessary inpatient days, which could lead to reduced overall costs.

* **Target Preventative Care:**

Since many admissions are from individuals aged 50–69 and those 70 and older, consider implementing targeted wellness and preventative care initiatives aimed at these age groups. This approach could facilitate early detection and management of chronic conditions, thereby decreasing emergency admissions.

* **Optimize Resource Allocation:**

With a higher volume of female patients and noticeable differences in length of stay across various races and ethnicities, it’s important to customize staffing and resource allocation (such as specialized nursing units and culturally competent care programs) to ensure efficient care and improve patient satisfaction.

* **Streamline Diagnosis and Procedure Protocols:**

Since osteoarthritis is the most frequently diagnosed condition and hip replacement is the leading procedure, standardizing best practices and care pathways for these issues could streamline care, potentially shorten length of stay, and enhance outcomes while lowering costs.

* **Feedback and Training:**

Utilize the collected data for regular performance evaluations and targeted training sessions to maintain consistency in care quality.

* **Target High-Risk Clinical Cases:**

Enhance Fracture Management: Given that fractures, particularly lower limb and hip fractures, carry a significantly higher mortality risk, focus on establishing specialized fracture care teams, especially for elderly patients.

* **Implement Bundled Payments:**

For high-volume procedures like joint replacements, adopting bundled payment models can improve cost predictability and encourage optimal care delivery.

**Forecasting**

I have utilized a linear forecasting model to project trends in yearly admissions and total costs over time. This approach relies on historical data to identify steady trends, making it effective for recognizing patterns and estimating future outcomes.

1. **Time Series Data Forecasted:**

* Admissions Count:This predicts future patient volumes, enabling hospitals to prepare for capacity needs.
* Total Costs:This estimates future expenditures, which is crucial for financial planning and resource allocation.

**How the Forecast Helps Answer Business Questions:**

Capacity Planning**:** It aids hospitals in preparing for potential fluctuations in patient volume.

Budget Forecasting**:** It helps in anticipating future costs and optimizing resource distribution.

Trend Identification**:** It reveals patterns in admissions and expenses, which supports long-term strategic planning.

1. **Forecast Accuracy:**

* Presence of prediction intervals offer a range of possible values, reflecting confidence in the estimates.
* Accuracy is influenced by historical consistency—external factors such as policy changes, pandemics, or economic shifts that can affect actual outcomes.
* The model was assessed using historical data to determine how accurately it predicted past values, ensuring a reasonable level of reliability.
* By utilizing linear forecasting, stakeholders can make informed decisions regarding resource management, financial planning, and operational efficiency.

**Calculated Fields Used**

1. **Operational KPIs (Efficiency & Resource Utilization):**

* **Length of Stay (LOS) (bin)** – Assesses how long patients stay in the hospital to evaluate overall efficiency.
* **Average LOS Per Hospital** – Monitors the typical duration of patient stays, providing insights into resource use.
* **Readmission Risk (High vs. Low)** – Analyzes the chances of patients returning for further treatment, which impacts hospital capacity planning.

1. **Financial KPIs (Revenue & Cost Management):**

* **Average Cost Per Hospital** – Determines the average expenditure for each hospital, aiding in budget management.
* **Average Cost Per Service Area** – Highlights cost differences across various healthcare service regions.
* **Cost-to-Charge Ratio** – Evaluates the relationship between hospital costs and billed charges, measuring financial performance.
* **Revenue Per Hospital** – Calculates the total income generated by each hospital, indicating its profitability.

1. **Patient Care KPIs (Quality & Outcomes):**

* **Mortality Risk (bin)** – Classifies patients according to their risk of mortality, assisting in risk evaluation.
* **Age Group (group)** – Segments patients by age to examine demographic patterns in hospital admissions.

**Visualization Techniques Used**

* **Severity (LOS & Costs) – Scatter Plot:** Illustrates the connection between length of stay (LOS) and costs, helping to spot patterns or anomalies.
* **Patient Disposition – Bar Chart:** Effectively compares the number of patients across different disposition categories (e.g., discharged, transferred, deceased).
* **Hospitals (LOS) - Bar Chart:** Shows the average length of stay in various hospitals for comparison.
* **Procedures (Costs) - Bar Chart:** Emphasizes cost differences among various procedures to pinpoint expensive treatments.
* **Admissions (Age Group & Gender-Wise) - Bar Chart:** Facilitates comparison of admissions across different age groups and genders.
* **Length Of Stay - Bar Chart**: Offers a clear view of how patient stays are distributed across various age groups and genders.
* **Admission (Race-Wise) – Pie Chart:** Displays the share of admissions by race to gain insights into demographic representation.
* **Admissions By State Level – Symbols Map:** Visually depicts geographic differences in hospital admissions across states.
* **Yearly Admissions & Costs Trend – Line Chart:** Monitors trends over time, illustrating how admissions and costs change annually.
* **Mortality Risk Assessment - Highlight Tables:** Utilizes color coding to swiftly identify areas with high mortality risk.
* **Readmission Risk - Highlight Tables:** Marks high-risk readmission groups to focus on necessary interventions.
* **Common Diagnoses vs. Length of Stay – Treemap:** Represents diagnoses and their related LOS in a compact format.
* **Cost-to-Charge Ratio – Text Table:** Offers exact numerical values for detailed cost analysis.
* **Distribution of Risk of Mortality Scores – Histogram:** Displays the frequency distribution of mortality risk scores for further examination.
* **Influence Of LOS On Hospital Charges- Scatter Plot:** Uncovers relationships between LOS and hospital charges, based on diagnosis.

**Interactivity and Dashboard Design**

* **Hover Tooltips:** Users can access extra information on data points without overcrowding the dashboard, which enhances clarity and insight.
* **Dynamic Highlighting:** This visually highlights critical areas for quick identification of important insights.
* **Clickable Charts (Selecting Age Group, Gender, Race to Filter Other Charts):** This boosts user engagement by enabling dynamic interaction with data.
* **Trend Forecasting Slider (Future Cost & Admissions Projections):** This aids decision-making by allowing users to simulate future scenarios.
* These interactive features improve usability, data exploration, and decision-making by making insights more accessible, actionable, and user focused.