



OPERATIONAL INSIGHTS FROM A HOSPITAL MANAGEMENT SYSTEM

Data Analytics Capstone Project

Author: Diego Pozo

Introduction

This capstone project uses a dataset from Kaggle's Hospital Management System to analyze hospital expenses from a departmental perspective. The data comes from a fully normalized, simulated hospital database that includes detailed information on appointments, bed usage, room admissions, staff roles, and billing. The goal is to explore how hospital resources are utilized across departments by examining total and average patient costs associated with appointments, beds, and rooms. Analyzing these costs at the department level can reveal patterns in resource use and help identify which departments are linked to higher patient-related expenses. The project will be carried out using Excel for initial preparation, SQL for data querying, and Tableau for visualizing the results.

1. Dataset Collection

1. SOURCE OF THE DATA

The dataset was obtained from Kaggle - [Hospital Management System](#). It is a simulated, fully normalized hospital database designed for educational and analytical purposes. Upon review, the dataset was found to be clean and well-structured. The tables were already normalized, column names were consistent, and there were no missing or duplicated values that required additional cleaning. Each table was simply converted from Excel format to .csv files in order to upload the data into SQLite for analysis.

2. GENERAL SUBJECT MATTER

The data represents healthcare operations, including information on patients, doctors, appointments, diagnoses, prescribed medications, and billing. It is structured to mimic a real hospital management system and provides a rich base for operational and financial analysis.

3. SIZE AND STRUCTURE

The dataset is composed of 14 relational tables containing a total of 93 columns, each with a different number of records depending on the type of data. For example, the "Appointments" table contains significantly more records than the "Doctors" table due to repeated patient visits. The dataset includes various data types such as numerical values (payment amounts), date and time fields (appointment dates), and categorical data (names, doct_id, nurse_id, helper_id, dept_id).

This project will primarily focus on three core service tables: Appointments, BedRecords, and RoomRecords. Each of these tables contains a billing-related field (payment_amount or amount) and is linked to one or more staff roles. These roles (such as doctors, nurses, and helpers) are connected to their respective departments via foreign keys. By following these relationships, it is possible to trace every patient's cost entry back to a specific department, allowing for a department-level cost analysis across services.

2. Analytical Process & Insights

1. SPECIFIC HYPOTHESES BEING TESTED WITH THE DATA

Analyzing total and average patient costs by department can uncover patterns in hospital resource use, helping identify which departments are associated with higher patient-related expenses.

2. PATTERNS AND INSIGHTS FROM THE ANALYSIS

The project seeks to identify patterns that may reveal strengths, weaknesses, or optimization opportunities in the hospital's operations:

- **Question 1: Which departments have the highest total and average costs per patient across all services (appointments, beds, and rooms)?**

To answer this question, I combined SQL and Tableau. I began by writing SQL queries to calculate the total and average cost per patient for each department based on data from appointments, bed records and room records. I created separate subqueries for each service type and linked them to department and staff tables so that each cost entry was correctly assigned. In each subquery I grouped the records by department and patient, summed the costs and counted distinct patient IDs (unique patients) to prevent double counting and ensure that the average cost truly reflected the number of individuals treated. Next, I merged all patient-level cost data into one table using UNION ALL and performed a final aggregation to produce a single row per department showing total cost, total unique patients and average cost per patient. I then exported this summary table and loaded it into Tableau for visualization.

In Tableau, I built a set of visualizations that allowed me to compare departments both in terms of total spending and cost efficiency per patient. I used a tree map to show the distribution of total costs by department and a bar chart to highlight departments with the highest average cost per patient. These visual tools made it easier to identify outliers and trends.

The results revealed that Critical Care was the department with the highest total patient costs, reaching over \$13.8 million, with an average cost per patient of \$64,197. General Medicine, although it served the largest number of patients, showed a much lower average cost of \$17,451 per patient. Surgery and Emergency Medicine also had high total costs, above \$8.5 million each, reflecting their role in delivering complex medical services. Interestingly, Psychology showed the highest average cost per patient, at \$99,411. However, this can be explained by the very low number of patient records in this department, suggesting that it is not commonly used or practiced within the hospital, which could contribute to the inflated cost per patient. These findings can help hospitals identify which areas are more resource-intensive and where there might be opportunities to improve cost-efficiency without compromising care quality.

- **Question 2: Do departments with more assigned staff roles (doctors, nurses, helpers) tend to generate higher costs?**

To address this question, I wrote SQL queries to count the number of doctors, nurses, and helpers assigned to each department. Each staff category was calculated separately, grouping by department. Then, I joined these counts into a single summary table alongside the total costs for each department, which were computed by summing expenses from appointments, bed stays, and room services. This comprehensive SQL query provided a unified view of staffing and cost data per department. After exporting the results to Tableau, I created a combined visualization that layered a bar chart representing total costs per department with a line chart showing total staff counts. This approach enabled a clear comparison of staffing levels against expenditures for each department.

The analysis showed a general trend where departments with more assigned staff tended to have higher total costs. For example, Critical Care had 136 staff and \$13.8 million in costs, General Medicine had 141 staff with \$9.4 million, and Surgery had 123 staff costing \$8.9 million. However, this trend is only partial since some departments exhibit high staffing numbers while maintaining relatively lower costs, as is the case with Family Medicine or the ones at the end of the line, indicating that maybe factors beyond staffing affect the cost fluctuation.

These findings suggest that while staffing is an important factor in cost generation, hospital administrators should also consider other operational variables when planning budgets and workforce allocations.

- **Question 3: Are there departments that achieve lower average patient costs while handling a similar number of patients?**

To explore which departments maintain lower average patient costs while managing a comparable patient load, I crafted a SQL query similar to the one used for the first question. The query aggregated total costs, and unique patient counts across all services, including appointments, bed records, and room records, combining these datasets using UNION ALL. Then, I summarized the data by department to calculate both total patient volume and average cost per patient. After preparing this summary, I imported it into Tableau for visualization.

In Tableau, I created a bar chart highlighting the top five departments that demonstrated the lowest average costs per patient while serving a significant number of patients. The departments that stood out were Cardiology, Neurology, Pulmonology, General Medicine, and Endocrinology. This visualization made it easy to compare cost efficiency across these high-volume departments, revealing patterns of resource management and potential operational best practices. By focusing on both patient volume and cost, this chart provides a balanced view that can help hospital management identify areas where efficiency is being achieved without sacrificing the quality or quantity of care. These findings offer insight into which departments operate with greater cost efficiency and may serve as models for cost management strategies within the hospital.

3. Business Recommendations

1. RECOMMENDATION 1

The hospital should conduct a detailed review of departments with unusually high average costs per patient, such as the department of Psychology for example. Given that Psychology serves a very small number of patients yet incurs exceptionally high costs, the hospital should evaluate whether maintaining this department as it currently operates is sustainable or if restructuring is necessary in this case. This assessment could reveal opportunities to improve efficiency or consider alternative service delivery models. Additionally, the hospital should focus on departments with high total costs driven by large patient volumes, like General Medicine and Cardiology. These departments significantly impact on the hospital's overall expenses and revenue, so optimizing resource use here is essential for financial stability and effective resource allocation.

2. RECOMMENDATION 2

The hospital should assess departments that maintain a high number of staff members but generate relatively low total patient costs. This trend, identified during the analysis for staffing vs total patient cost by department, raises potential concerns about operational efficiency. When a department has substantial staffing but contributes minimally to the overall cost structure, it may be a sign of low patient throughput or limited demand for its services. This does not necessarily imply poor performance but rather highlights an opportunity to investigate whether staff resources are being used effectively. By reviewing these departments more closely, the hospital can make informed decisions about redistributing staff, optimizing schedules, or even redefining departmental roles to better align resource allocation with actual service needs and patient impact.

4. Next Steps

As a future direction, the hospital should consider incorporating clinical outcome data such as patient recovery rates, length of stay, or readmission rates to complement the current cost-based analysis. This would allow for a better understanding of whether departments with higher costs are also delivering greater value in terms of patient outcomes. For example, some departments may appear expensive, but if they are managing complex cases with strong recovery results, their costs may be justified.

Additionally, gaining access to more detailed information such as the specific procedures each patient underwent, staff working hours, and the usage of medical resources would allow for more precise cost tracking. With this level of data, it would be possible to link each patient to the specific departments and services they used. This would make the analysis more accurate and actionable while also helping identify the true sources of inefficiencies and guiding more focused decisions around resource allocation, staffing, and service design.