Healthcare Financial Analysis

Executive Summary

55,500	$1.42B	5 Years	15.5 Days
Patient Records	Total Revenue	May 2019-Jun 2024	Avg Length of Stay

Research Questions & Findings

1. Does insurance provider correlate with billing amounts?

Finding: No meaningful correlation. All five major insurers (Medicare, Blue Cross, Aetna, Cigna, UnitedHealthcare) show average billing within $227 of each other ($25,389-$25,616), representing less than 1% variance. This demonstrates highly standardized pricing regardless of payer.

2. Are certain medical conditions more expensive to treat?

Finding: Minimal variation across conditions. Obesity ($25,806) to Cancer ($25,162) shows only $644 difference (2.5% variance). All six conditions cluster tightly around $25,500, with uniform cost variability (54-56% coefficient of variation) reflecting synthetic data generation rather than real-world clinical complexity.

3. Which admission type has the longest hospital stays?

Finding: Virtually identical across all types. Emergency (15.60 days), Elective (15.53 days), and Urgent (15.41 days) differ by only 0.19 days or 4.5 hours. Urgent admissions show highest efficiency at $1,656 revenue per day, while Emergency generates $1,635/day despite slightly longer stays.

Key Metrics Summary

Category	Highest	Lowest
Insurance Billing	Medicare: $25,616	UnitedHealth: $25,389
Condition Costs	Obesity: $25,806	Cancer: $25,162
Length of Stay	Emergency: 15.60 days	Urgent: 15.41 days
Revenue per Day	Urgent: $1,656/day	Emergency: $1,635/day

Critical Analysis: Synthetic Data Patterns

This dataset, generated using Python's Faker library for educational purposes, exhibits remarkable uniformity across all dimensions, a characteristic that distinguishes synthetic from real-world healthcare data:

- **Insurance parity:** 0.9% variance vs. real-world 15-30% differences due to negotiated rates

- **Condition costs:** 2.5% variance vs. real-world 50-200% (cancer typically 3-5x more expensive than routine care)
- **Length of stay:** 0.2 days difference vs. real-world 5-15 days (emergency typically 2-3x longer than elective)
- **Distribution uniformity:** Perfect 20%/33% splits across categories vs. real-world skewed distributions

Value Proposition: While this synthetic dataset does not reveal actionable cost optimization opportunities (due to its uniformity), it provides an ideal environment for demonstrating SQL proficiency, database design, and analytical methodology. The techniques applied here (CTEs, window functions, multi-dimensional aggregation, statistical calculations) transfer directly to messy, complex real-world healthcare data where they generate genuine business value.

SQL Techniques Demonstrated

- Common Table Expressions (CTEs) for multi-step calculations
- Window functions (NTILE, RANK, LAG) for percentile and trend analysis
- Date arithmetic for length of stay and temporal analysis
- Statistical aggregations (STDDEV, coefficient of variation)
- Multi-dimensional GROUP BY for cross-tabulation
- Conditional aggregation and MODE for profile analysis

Applicability to Real Healthcare Data

In production healthcare analytics environments, these same SQL techniques would reveal:

- Significant cost optimization opportunities (high-variance conditions)
- Resource allocation inefficiencies (emergency department bottlenecks)
- Insurance contract negotiation leverage points (payer-specific pricing patterns)
- Clinical pathway standardization targets (high-cost outlier cases)
- Readmission reduction strategies (patient segmentation by risk factors)

Conclusion: This analysis demonstrates proficiency in healthcare data analysis, SQL querying, and critical thinking about data quality and applicability. These skills transfer directly to operational healthcare analytics roles.