Healthcare Data Analysis

SQL Insights on Hospital Operations

Amirhossein Tonekaboni

ERP Data Analyst
SAP Business One Consultant

🔗 ATonekaboni.github.io
in Linkedin.com/in/Tonekaboni

Healthcare Data Analysis

SQL-based investigation of Massachusetts Hospital operations for strategic decision-making and annual reporting.

Project Overview

This analysis examines over 10 years of patient data from Massachusetts General Hospital (2011-2022), covering approximately 1,000 patient records with medical encounters, treatment costs, insurance coverage, and medical procedures. The investigation focuses on identifying revenue leakage from insurance coverage gaps, understanding patient readmission patterns, and analyzing encounter distribution across different care types.

The dataset provides insights into operational challenges facing healthcare administrators who must balance cost containment with quality care delivery. By analyzing historical patterns, this investigation reveals actionable findings that support evidence-based decision making for strategic planning and operational improvements.

Database Architecture

The healthcare database connects patient demographics to clinical encounters and financial records through four normalized tables: **patients**, **encounters**, **procedures**, and **payers**. This structure enables multi-dimensional analysis across demographics, encounter types, procedure patterns, and financial performance.
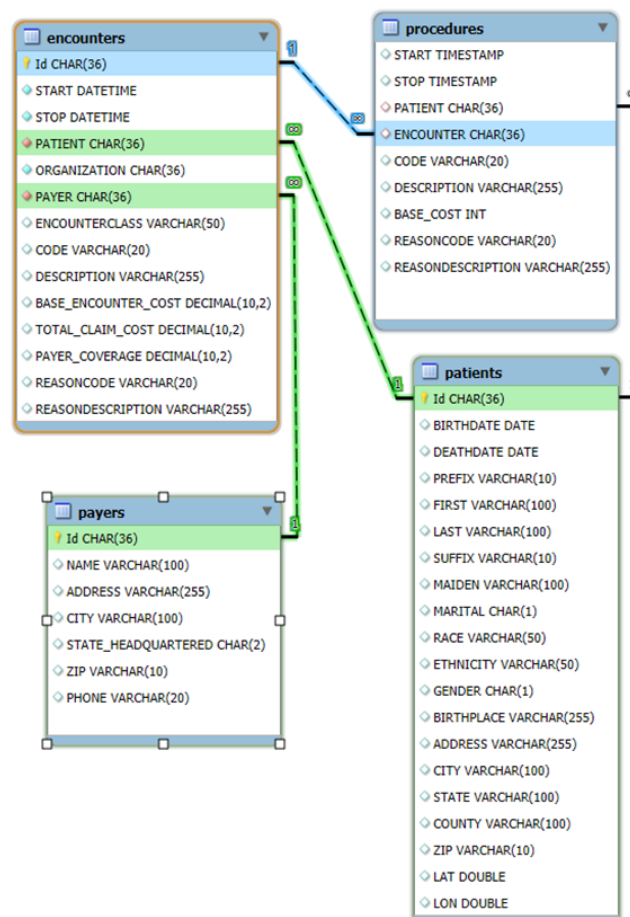
Repository Structure

This repository contains the complete healthcare analytics project with the following components:

- **hospital_analytics.sql** - Complete SQL analysis script implementing three analytical objectives
- **dataset.rar** - Massachusetts General Hospital synthetic data (2011-2022) - database file
- **README.md** - Comprehensive project documentation with findings and recommendations
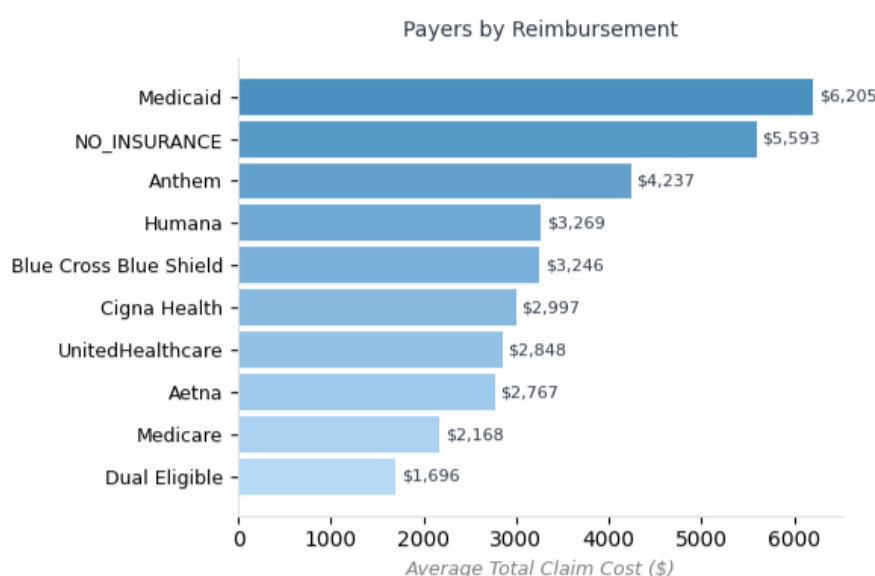
Executive Summary

Massachusetts General Hospital analysis reveals critical revenue risks with 49% of encounters lacking payer coverage, representing potential losses exceeding $75 million annually. Patient volume grew 164% over the decade, but 2021 data shows concerning readmission patterns with 771 patients requiring readmission within 30 days. The analysis identifies actionable opportunities for revenue recovery through pre-visit insurance verification and targeted case management for high-risk patients.

Key Findings

Revenue Coverage Crisis

- **payer coverage:** 49% of all encounters (13,586 visits) operate without payer coverage, representing substantial revenue exposure affecting nearly half of hospital activity.

- **Reimbursement:** Payer reimbursement varies dramatically, with Medicaid providing the highest average at $6,205 per encounter, while Dual Eligible patients generate the lowest reimbursement at only $1,696 per encounter. Uninsured patients generate $5,593 in average claims, indicating significant uncollectable receivables.

  - *Visualization created in Python with SQLAlchemy and Matplotlib.*

Payers by Reimbursement

Payer	Average Total Claim Cost ($)
Medicaid	$6,205
NO_INSURANCE	$5,593
Anthem	$4,237
Humana	$3,269
Blue Cross Blue Shield	$3,246
Cigna Health	$2,997
UnitedHealthcare	$2,848
Aetna	$2,767
Medicare	$2,168
Dual Eligible	$1,696

Critical Care Economics

The analysis reveals dramatic cost variations in hospital procedures. ICU admissions dominate at $206,260 per procedure (5 total), representing the most resource-intensive care. However, the most striking finding is electrical cardioversion with 1,383 procedures at $25,903 each (over $35 million) - indicating substantial resource allocation to cardiac emergency procedures.

CODE	DESCRIPTION	avg_base_cost	total_procedures
305351004	Admit to ICU (procedure)	206260.4000	5
232717009	Coronary artery bypass grafting	47085.8889	9
392021009	Lumpectomy of breast (procedure)	29353.0000	5
302497006	Hemodialysis (procedure)	29299.5556	27
447365002	Insertion of biventricular implantable cardioverter defib...	27201.0000	4
180325003	Electrical cardioversion	25903.1106	1383
43075005	Partial resection of colon	25229.2857	7
432231006	Fine needle aspiration biopsy of lung (procedure)	23141.0000	1
433112001	Percutaneous mechanical thrombectomy of portal vein ...	20228.0351	57
415070008	Percutaneous coronary intervention	19728.0000	9

Average Cost

In contrast, high-volume procedures demonstrate standardized operational pricing across routine hospital services. The majority of frequent procedures operate at a consistent $431 cost structure - health assessments (4,596 procedures), hospice care (4,098), depression screening (3,614) all follow this uniform rate. However, renal dialysis emerges as a significant exception with 2,746 procedures at $1,004 each, representing nearly 2.5 times the standard rate and highlighting the resource-intensive nature of chronic kidney disease management.

CODE	DESCRIPTION	total_procedures	avg_base_cost
710824005	Assessment of health and social care needs (procedure)	4596	431.0000
385763009	Hospice care (regime/therapy)	4098	431.0000
171207006	Depression screening (procedure)	3614	431.0000
454711000124102	Depression screening using Patient Health Questionnai...	3614	431.0000
428211000124100	Assessment of substance use (procedure)	2906	431.0000
265764009	Renal dialysis (procedure)	2746	1004.0863
762993000	Assessment using Morse Fall Scale (procedure)	2422	431.0000
710841007	Assessment of anxiety (procedure)	2288	431.0000
430193006	Medication Reconciliation (procedure)	2284	509.1243
713106006	Screening for drug abuse (procedure)	1484	431.0000

high-volume procedures

Patient Safety Concerns

771 patients required readmission within 30 days. The most extreme case shows 1,376 readmissions for a single patient, representing both safety risks and resource consumption.

FIRST	LAST	num_readmissions
Kimberly627	Collier206	1376
Mariano761	OKon634	876
Shani239	Parisian75	871
Gail741	Glover433	442
Ward668	Nicolas769	421

Readmission

```sql
1  -- Top 5 patients with most readmissions (with patient names)
2  WITH person AS (
3      SELECT e.START, e.STOP, p.Id, p.FIRST, p.LAST
4      FROM encounters e
5      JOIN patients p ON e.PATIENT = p.Id
6  ),
7  next AS (
8      SELECT   *, LEAD(START) OVER(PARTITION BY Id ORDER BY START) AS next_start_date
9      FROM person
10 )
11 SELECT Id, FIRST, LAST, COUNT(*) AS num_readmissions
12 FROM next
13 WHERE DATEDIFF(next_start_date, STOP) < 30
14 GROUP BY Id, FIRST, LAST
15 ORDER BY num_readmissions DESC
16 LIMIT 5;
```

SQL script for calculating patients with highest number of readmissions

Operational Efficiency

Encounter Volume and Patient Flow Analysis: Hospital encounters grew from 1,336 in 2011 to peak at 3,530 in 2021 - a 164% increase over ten years. The quarterly unique patient data reveals that 2021's record encounter volume coincided with dramatic patient spikes in Q1-Q2 (417 and 414 unique patients respectively) before declining to 279 and 223 in Q3-Q4. This pattern shows that 2021's high encounter volume resulted from both increased patient volume and higher visit frequency per patient. (The 2022 data included only first quarter of the year.)

YR	TOTAL_ENCOUNTERS	yr	qtr	total_unique_patient
2011	1336	2019	1	229
2012	2106	2019	2	231
2013	2495	2019	3	226
2014	3885	2019	4	232
2015	2469	2020	1	229
2016	2451	2020	2	235
2017	2360	2020	3	249
2018	2292	2020	4	258
2019	2228	2021	1	417
2020	2519	2021	2	414
2021	3530	2021	3	279
2022	220	2021	4	223
		2022	1	103

                Total Encounters   *Unique Patients*

Encounter Duration Performance confirms exceptional operational efficiency with 96% of encounters completing within 24 hours, indicating effective patient throughput management and discharge planning protocols across all service lines.

Service Line Distribution: Ambulatory care dominated hospital operations, ranging from 36.9% (2021) to 60.3% (2014), while outpatient services spiked to 40.2% in 2021 - suggesting temporary operational shifts. Urgent care increased dramatically from 2.2% (2011) to 17.8% (2019), while emergency care remained stable between 4-11%. Wellness care declined from 13.0% (2011) to 4.4% (2021), and inpatient care consistently declined from 6.2% to 1.6%, reflecting the shift toward ambulatory treatment.

YR	ambulatory	outpatient	wellness	urgentcare	emergency	inpatient
2011	49.9	24.5	13.0	2.2	4.1	6.2
2012	42.5	21.1	9.1	14.2	8.7	4.4
2013	44.3	19.4	7.5	14.4	9.0	5.4
2014	60.3	17.9	4.9	8.4	5.6	3.0
2015	43.5	20.5	6.9	15.4	9.2	4.5
2016	43.8	19.6	7.4	13.9	10.2	5.1
2017	41.8	20.1	7.2	16.3	9.2	5.3
2018	40.7	20.9	7.6	16.4	10.8	3.5
2019	38.0	20.5	7.5	17.8	10.2	6.0
2020	47.3	19.7	6.3	14.5	9.3	2.9
2021	36.9	40.2	4.4	10.7	6.3	1.6
2022	55.0	16.8	5.5	12.7	8.2	1.8

Service Line Distribution

Strategic Recommendations

Revenue Leakage With 49% of encounters having no insurance coverage, implement automatic insurance checks 24-48 hours before patient visits. This could recover millions in lost revenue and should be the top priority given how many patients are currently uninsured.

Target High-Risk Patients Deploy case management for the 771 patients with 30-day readmissions, especially targeting the top 5 patients averaging nearly 800 readmissions each. Better care coordination could significantly reduce both costs and safety risks while improving patient outcomes.

Improve Insurance Payment Strategy Dual Eligible patients pay only $1,696 per encounter while Medicaid pays $6,205 - nearly four times more than the lowest payer. The hospital should focus on attracting patients with better-paying insurance plans and negotiate higher rates with the lowest-paying insurers like Dual Eligible.

Technical Implementation

- **Database:** MySQL with four-table normalized design connecting patient demographics to clinical encounters and financial records.

- **Core Analysis Methods:** LEAD window function to track patient readmission intervals, CASE statement aggregation for encounter type distribution, and multi-table joins linking procedures to cost and coverage data.

- **Dataset:** 11 years of operational data (2011-2022) covering 1,000+ patient records and 27,000+ encounters sourced from Maven Analytics Hospital Patient Records.

Project Limitations

- **Incomplete 2022 data:** Only first quarter available, preventing full-year trend comparison.
- **Readmission Analysis:** 30-day window follows standard clinical practice but may miss longer-term patterns.
- **Cost interpretation:** Analysis uses hospital base costs, not actual patient bills or final insurance payments.
- **Scope:** Single hospital system data may not represent national healthcare patterns or rural facilities.

Prepared by

Amirhossein Tonekaboni

ERP Data Analyst
SAP Business One Consultant

- GitHub Repository
- Portfolio
- LinkedIn