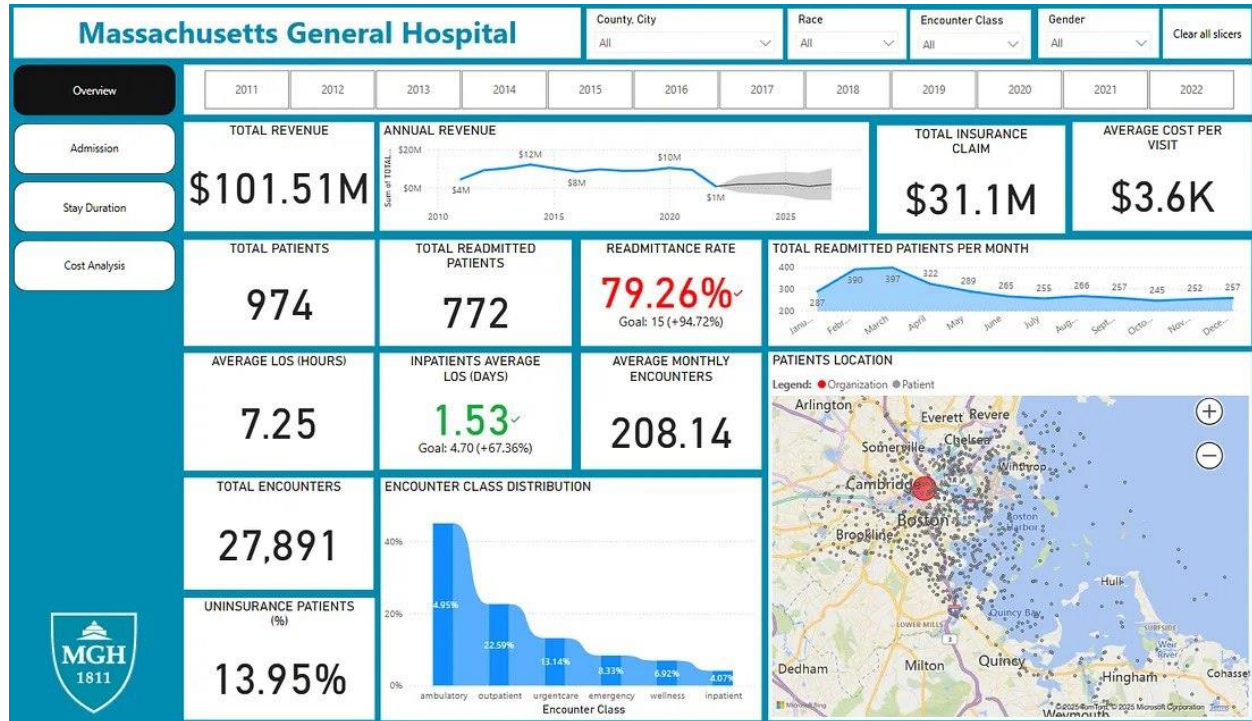


A Decision Support System for Optimizing Healthcare Operations and Patient Outcomes

Tools Used: Python, Power BI, SQL



Key Business Needs

Healthcare organizations generate large volumes of data that, when effectively analyzed, can significantly improve patient outcomes and optimize operations. However, fragmented data sources often create silos, limiting comprehensive analysis and hindering timely decision-making.

This project developed a Decision Support System (DSS) that empowers healthcare administrators with insights to improve patient care, streamline operations, and strengthen financial sustainability.

Data Sources

The project utilized publicly available healthcare datasets, including:

- Encounters (27,891 records, 14 fields)
- Patients (974 records, 20 fields)

- Payers (10 records, 7 fields)
- Procedures (47,701 records, 9 fields)
- Organizations (1 record, 8 fields)

These datasets provided the foundation for analyzing patient demographics, encounters, costs, payer information, and operational trends.

Key Performance Indicators (KPIs)

The DSS tracked performance across four core areas:

- Operational Efficiency: Average length of stay, encounter class distribution
- Financial Management: Payer coverage percentage, average cost per encounter
- Geographical Access: Patient proximity and encounter distribution by location
- Care Quality: Readmission rates and patient outcomes

These KPIs provided a comprehensive lens to evaluate both operational and clinical performance.

Data Preparation

Duplicate Checks

- No duplicate records were found across patient, payer, or encounter datasets.

```
#checking patient dataset shape before dropping any duplicates
patientsdf.shape

(974, 20)
```

```
#drop any duplicates (row) from the patients dataset
patientsdf.drop_duplicates(inplace=True)
patientsdf.shape

(974, 20)
```

Missing Values

- Columns with >70% missing values were dropped.
- Non-critical fields were excluded to preserve dataset integrity.
- Missing payer data was filled with “N/A” to represent uninsured patients.

```
#checking for missing values
payers_missing = (payersdf.isna().mean() * 100).round(0).astype(str) + '%'
print(payers_missing)
```

```
Id          0.0%
NAME        0.0%
ADDRESS     10.0%
CITY        10.0%
STATE_HEADQUARTERED 10.0%
ZIP         10.0%
PHONE       10.0%
dtype: object
```

```
# Fill missing values with 'N/A'
payersdf.fillna('N/A', inplace=True)

# Check if the missing values have been filled
payers_missing = (payersdf.isna().mean() * 100).round(0).astype(str) + '%'
print(payers_missing)
```

```
Id          0.0%
NAME        0.0%
ADDRESS     0.0%
CITY        0.0%
STATE_HEADQUARTERED 0.0%
ZIP         0.0%
PHONE       0.0%
dtype: object
```

Outlier Analysis

- Encounter costs showed expected variability (major surgeries, specialized treatments).
- Outliers in base costs were retained as valid.
- All patient demographic values fell within reasonable ranges.

```
encounterdf.describe()
```

| | CODE | BASE_ENCOUNTER_COST | TOTAL_CLAIM_COST | PAYER_COVERAGE |
|-------|--------------|---------------------|------------------|----------------|
| count | 2.789100e+04 | 27891.000000 | 27891.000000 | 27891.000000 |
| mean | 2.972670e+08 | 116.181614 | 3639.682174 | 1114.965652 |
| std | 2.017839e+08 | 28.410082 | 9205.595748 | 4768.615576 |
| min | 1.505002e+06 | 85.550000 | 0.000000 | 0.000000 |
| 25% | 1.853450e+08 | 85.550000 | 142.580000 | 0.000000 |
| 50% | 1.853490e+08 | 136.800000 | 278.580000 | 28.440000 |
| 75% | 4.244410e+08 | 142.580000 | 1412.530000 | 155.770000 |
| max | 7.029270e+08 | 146.180000 | 641882.700000 | 247751.420000 |

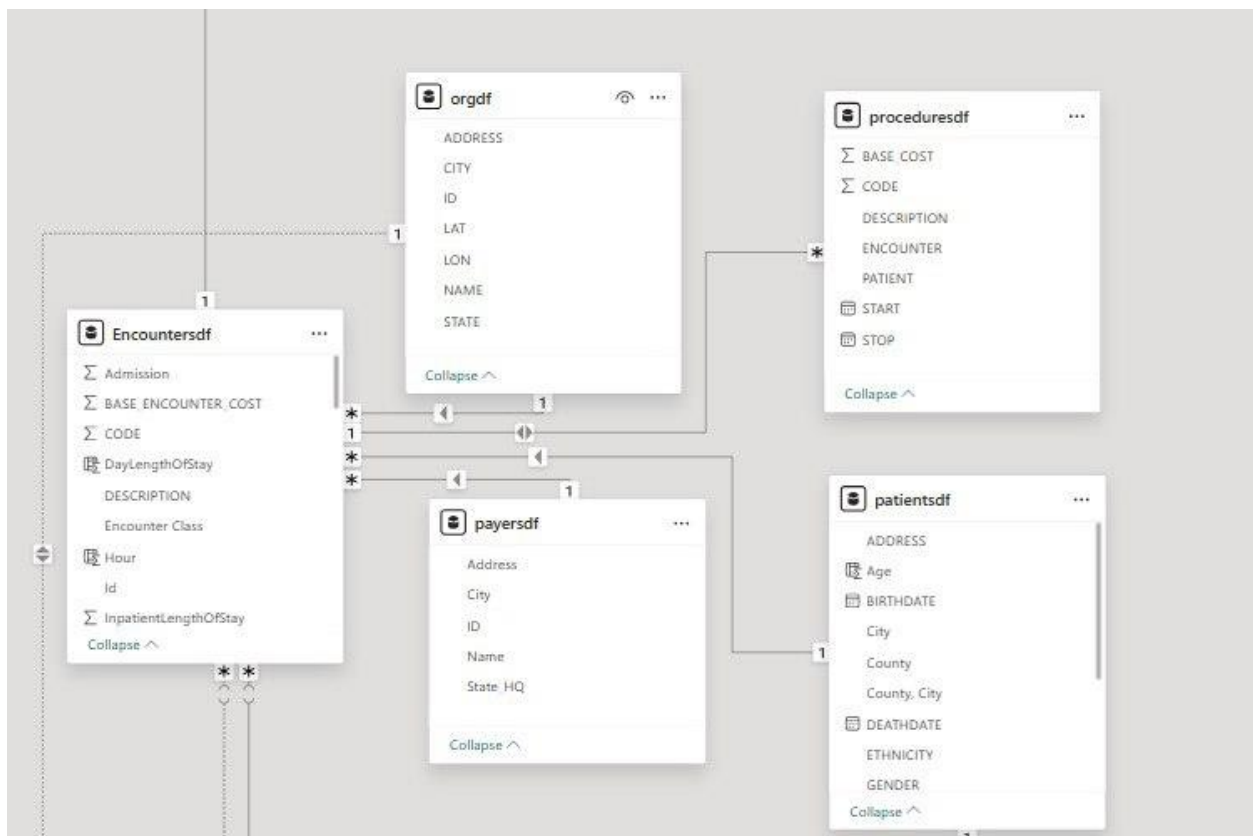
Data Modeling

A relational model was built linking:

- Encounters → Patients (one-to-many)

- Procedures → Encounters (many-to-one)
- Payers → Patients (one-to-one)
- Organizations → Encounters (contextual reference)

This structure enabled cross-dataset analysis of patient journeys, payer coverage, and healthcare procedures.



Dashboards & Measures

Power BI dashboards were created with custom DAX measures, including:

- Average Length of Stay
- Total Revenue
- Readmission Rate
- Uninsured Patient Rate

- Average Monthly Encounters

These visualizations provided stakeholders with an interactive, real-time view of healthcare operations.



Insights & Recommendations

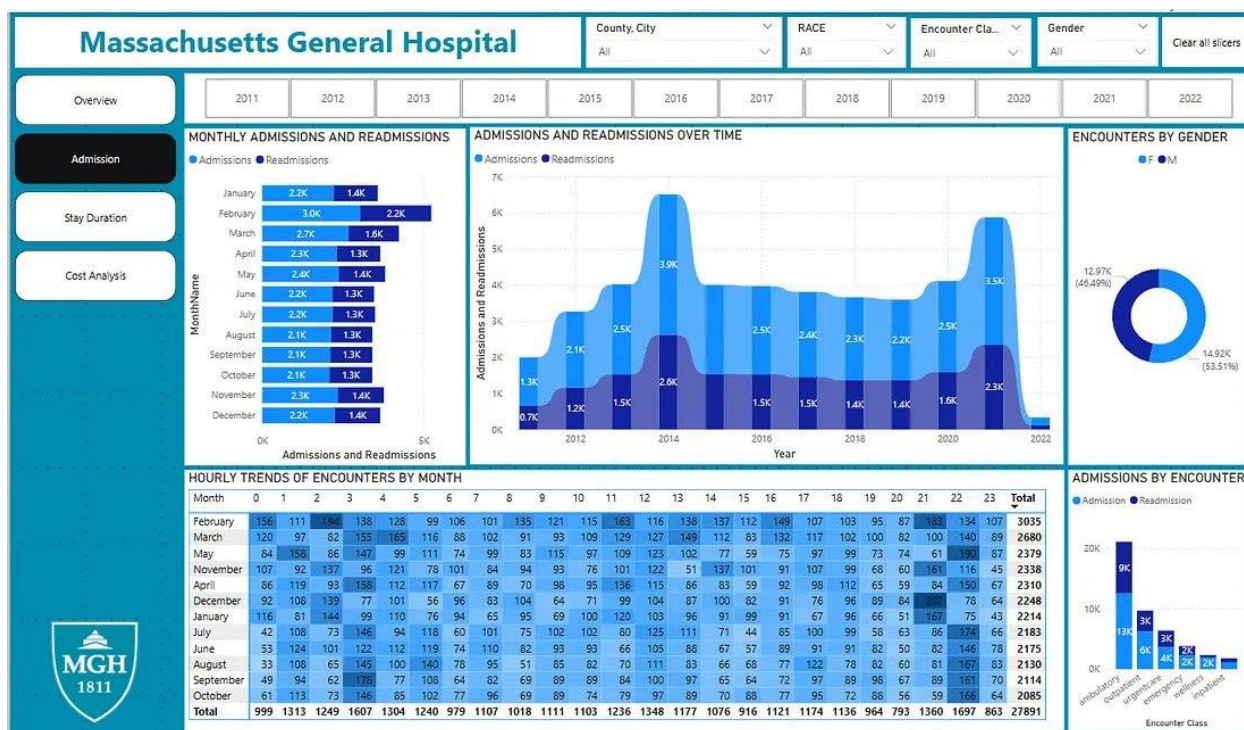
- Total Revenue: \$101.5M
- Average Cost per Visit: \$3.6K
- Total Insurance Coverage: \$31.1M
- Total Patients: 974

- Average Length of Stay: 7.25 hours
- Readmission Rate: 79.26%
- Average Monthly Encounters: 208

Notable Trends

- Encounters were highest in Boston, followed by Weymouth.
- Admissions peaked in 2014, declined, then rose again in 2021.
- Ambulatory encounters (13K) were the most common, followed by outpatient (6K).
- Gender split: 53.5% female, 46.5% male.
- Readmission rate of 79.26% highlights gaps in post-discharge care.

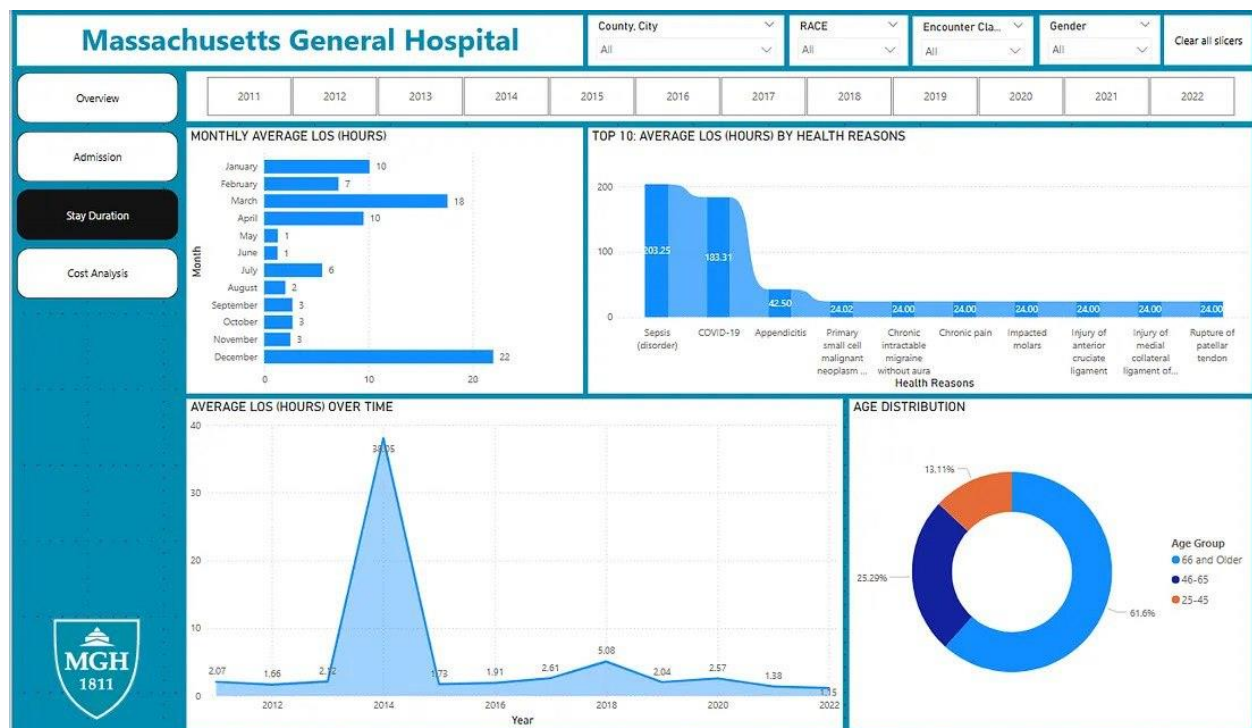
Appendix (Power BI Pages)



- Admissions Dashboard (Admissions & Readmissions)



• Cost Analysis Dashboard (Claim Cost Distribution)



• Stay Duration Dashboard (Length of Stay)