Healthcare Analytics System
Comprehensive Dataset Report

Generated: December 20, 2025

1. Executive Summary

This report provides comprehensive documentation of the healthcare analytics dataset ecosystem, consisting of five interconnected CSV files capturing patient demographics, physician performance, departmental operations, financial metrics, and physician registry information spanning January 2022 through December 2024. The primary objective is to enable patient readmission risk prediction through machine learning while supporting descriptive analytics for operational insights.

Dataset	Records	Columns	Purpose
patient_demographics.csv	1,001	7	Patient characteristics & outcomes
physician_performance.csv	3,960	10	Monthly physician metrics
department_metrics.csv	612	10	Department operational data
financial_performance.csv	36	10	Hospital-wide financials
physician_registry.csv	110	7	Physician directory

2. Patient Demographics Analysis

The patient demographics dataset contains 1,001 aggregated records representing patient groups by age, gender, and insurance type. Key features include patient count, average length of stay, average treatment cost, and readmission rate.

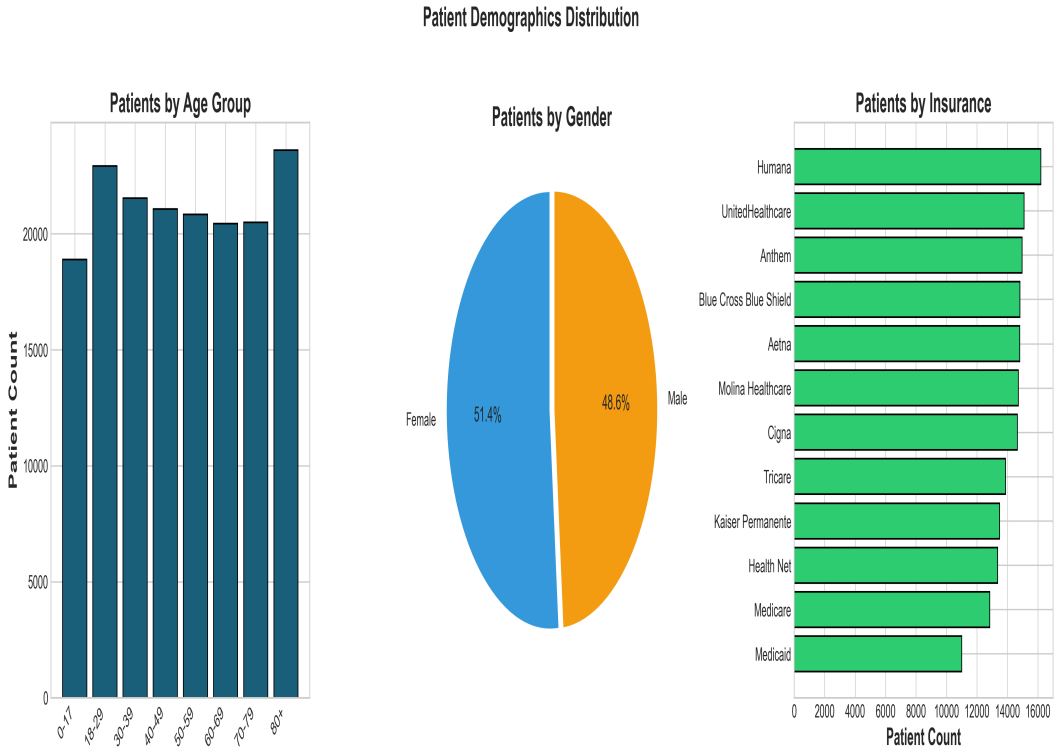

Figure 1: Patient Demographics Distribution

3. Readmission Rate Analysis

Readmission rates were analyzed across demographic segments to identify high-risk populations. A threshold of 20% was established as the boundary for high-risk classification, enabling targeted intervention strategies.

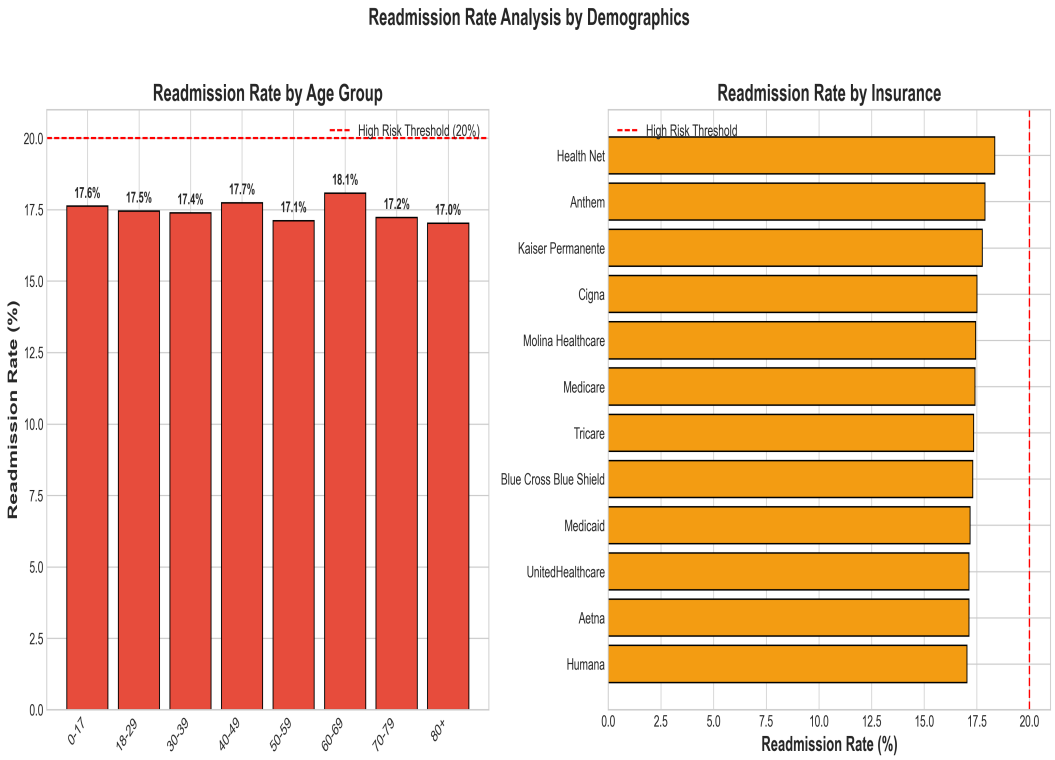

Figure 2: Readmission Rates by Demographics

4. Physician Performance Trends

Analysis of 110 physicians across 36 months reveals performance patterns including patient satisfaction scores and volume trends. Top-performing physicians demonstrate consistent satisfaction scores above 4.0 on a 5-point scale.

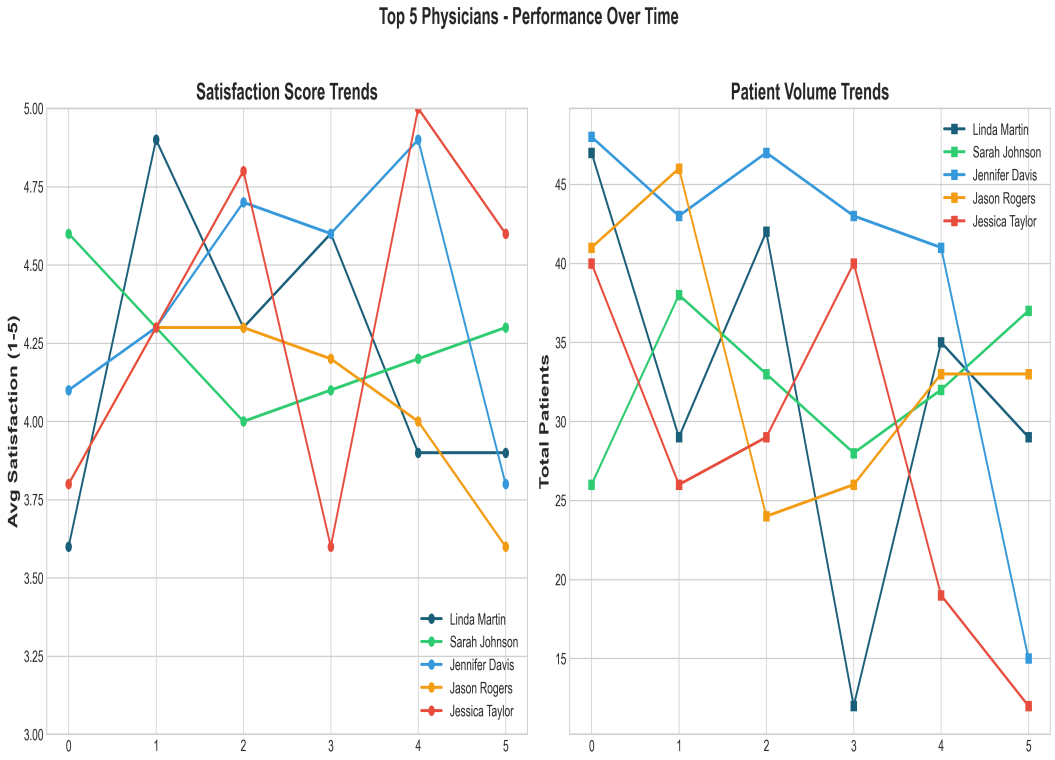

Figure 3: Top 5 Physicians Performance Trends

5. Department Metrics Comparison

Department-level analysis across 17 clinical departments shows variations in admissions, costs, and occupancy rates. Emergency Medicine leads in total admissions while specialty departments show higher per-patient costs.

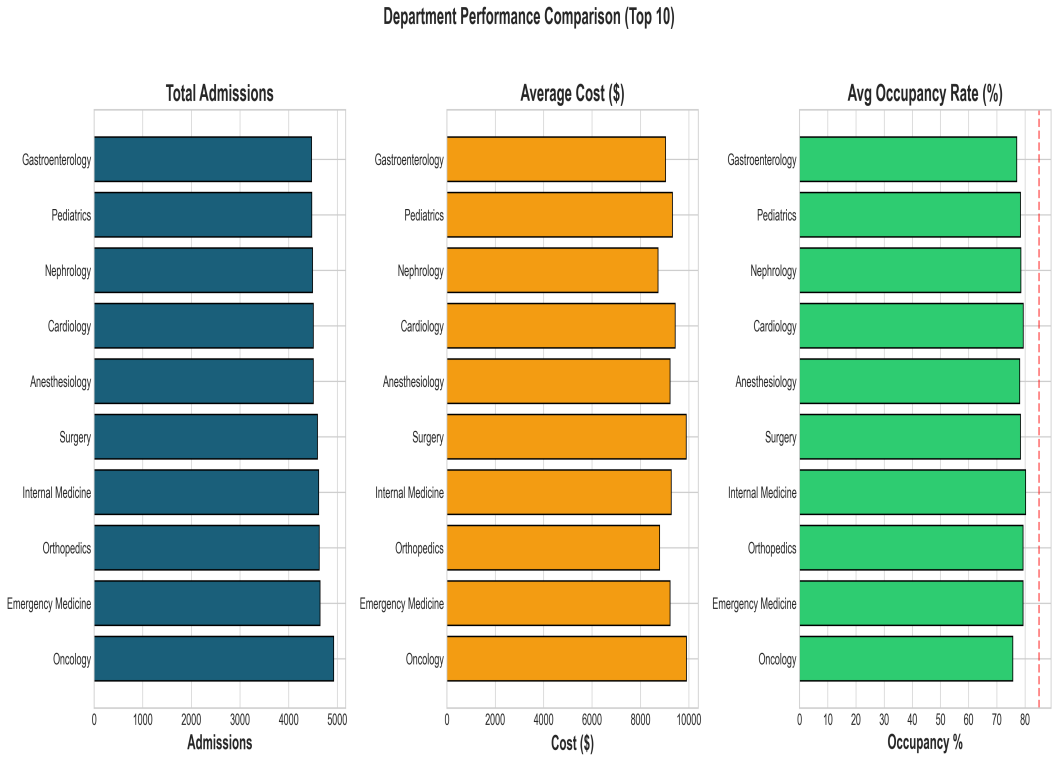

Figure 4: Department Performance Comparison

6. Financial Performance Dashboard

Financial analysis reveals consistent positive operating margins averaging 18.8% across the analysis period, with total revenue of approximately $709 million over three years and net income of $134 million.

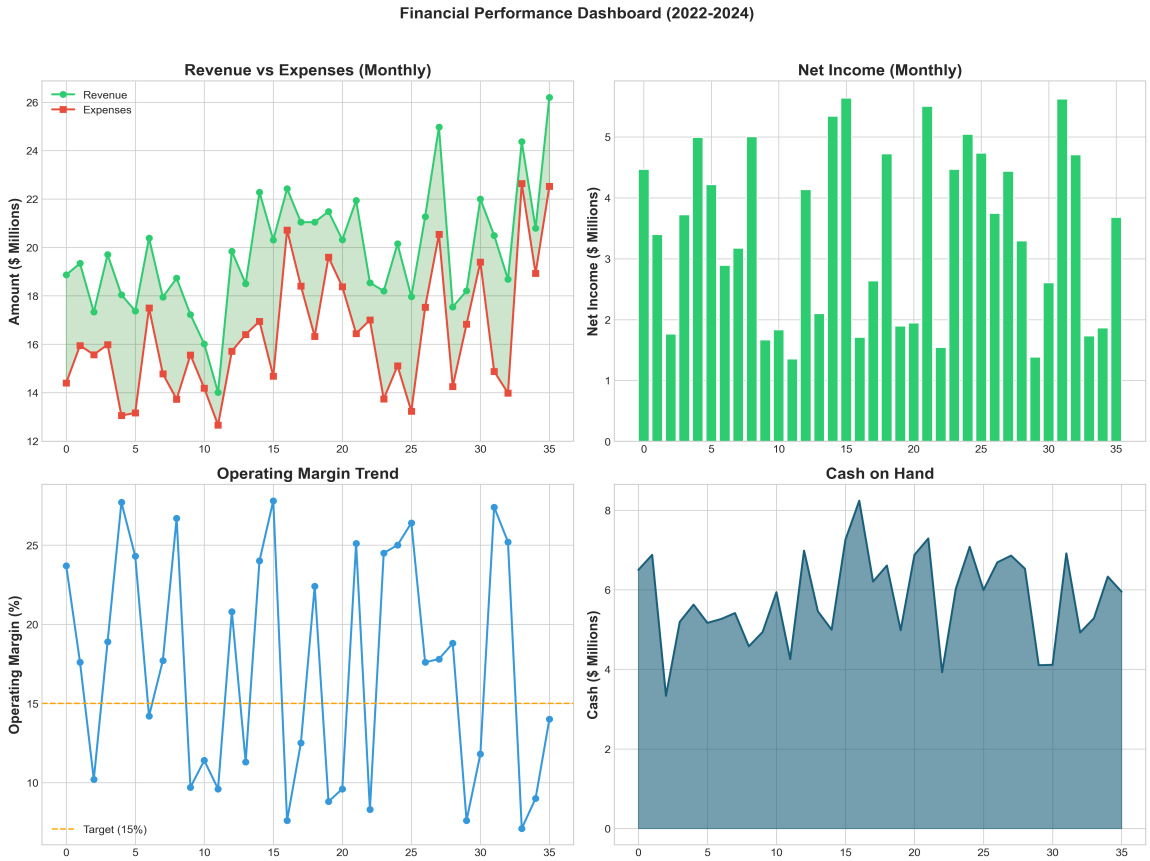

Figure 5: Financial Performance 2022-2024

7. Healthcare Cost Analysis

Cost analysis by insurance type and age group reveals significant variations. Understanding these patterns supports resource allocation and pricing strategies.

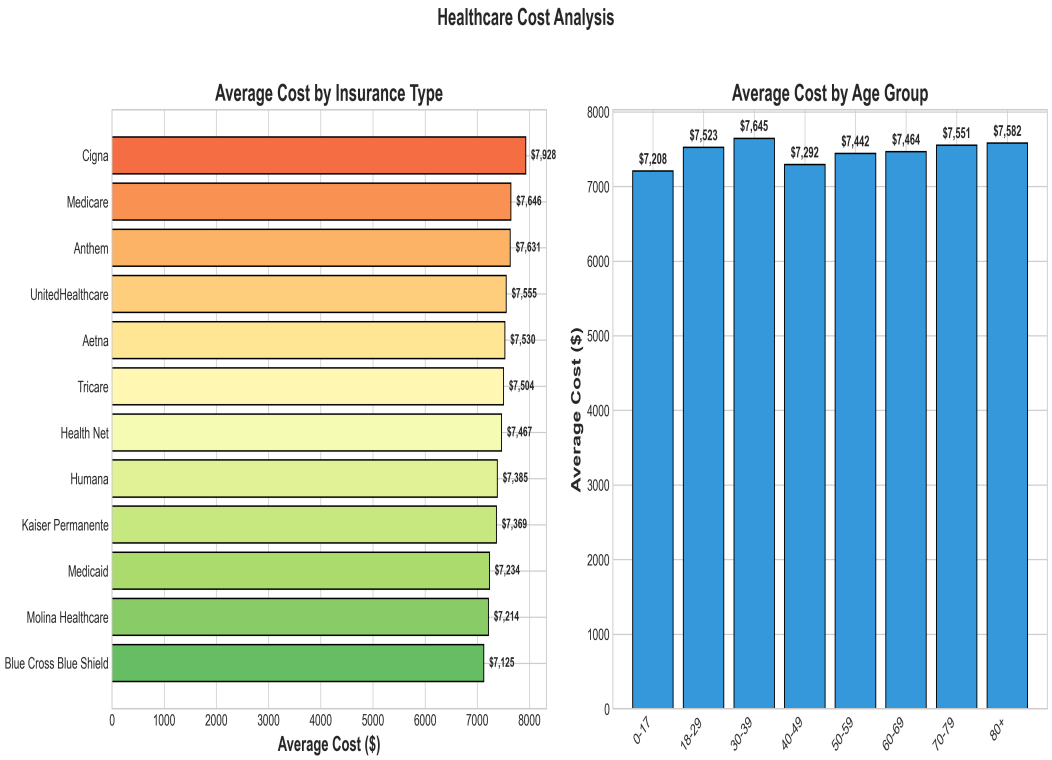

Figure 6: Cost Analysis by Demographics

8. Feature Correlation Analysis

Correlation analysis of numerical features reveals relationships between patient count, length of stay, cost, and readmission rate. These correlations inform feature selection for machine learning models.

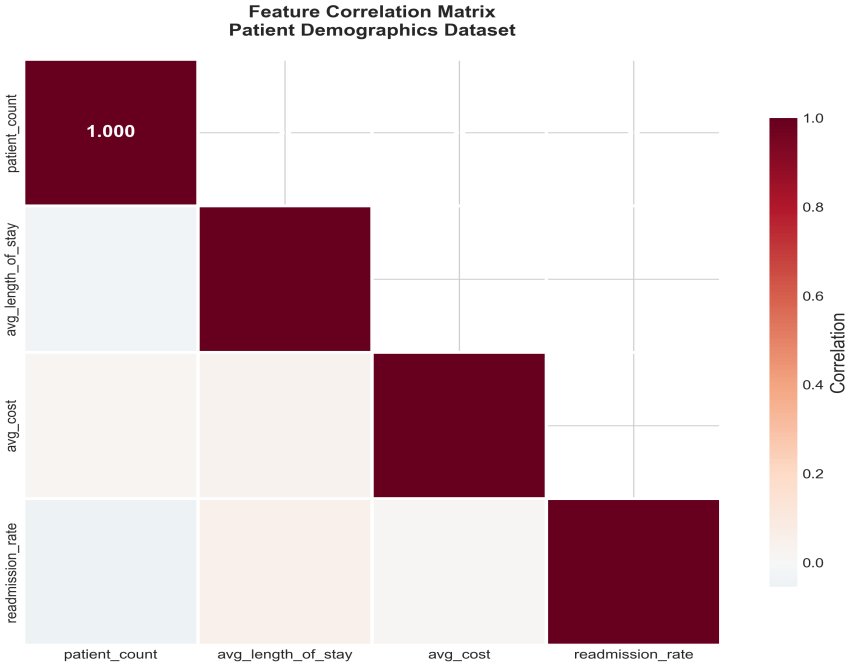

Figure 7: Feature Correlation Matrix

9. Conclusion

This comprehensive dataset infrastructure supports the healthcare system's strategic goal of reducing preventable readmissions through data-driven patient identification. The five datasets provide complete coverage of patient, physician, department, and financial dimensions, enabling both descriptive analytics and predictive modeling applications.