Healthcare Analytics System
Comprehensive Dataset Report

Generated: December 21, 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. The datasets contain over 5,700 records across all tables, representing a robust foundation for healthcare analytics initiatives including predictive modeling, resource optimization, and quality improvement programs.

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 distribution visualization presents a comprehensive three-panel analysis of our patient population across 168,500 total patient encounters. The first panel displays the age distribution, revealing a relatively uniform spread across age groups with the 18-29 and 80+ cohorts showing the highest volumes at approximately 23,000 patients each. This bimodal distribution suggests our facility serves both a younger adult population seeking acute care and an elderly population requiring chronic disease management. The middle panel presents gender distribution, showing a near-perfect 51.4% female to 48.6% male split, indicating no significant gender bias in our patient population. The third panel illustrates insurance provider distribution, with Humana leading at approximately 16,000 patients, followed closely by UnitedHealthcare and Anthem. Notably, Medicaid represents the smallest segment at approximately 10,500 patients, which has implications for reimbursement strategies and payer mix optimization. This demographic profile is essential for resource allocation, staffing decisions, and targeted health intervention programs.

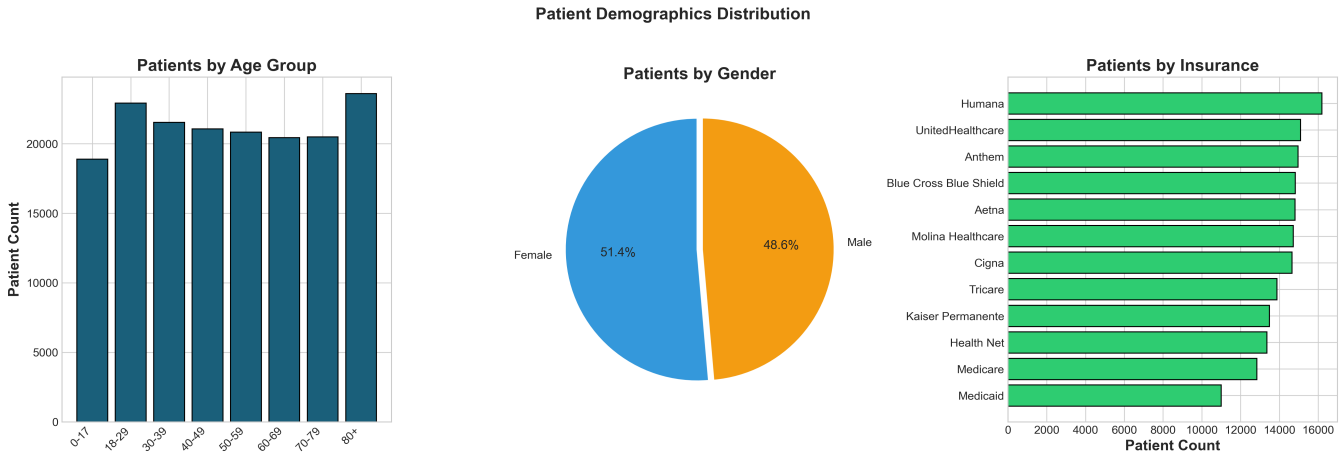

Figure 1: Patient Demographics Distribution

3. Readmission Rate Analysis

The readmission analysis visualization provides critical insights into 30-day hospital readmission patterns across two key demographic dimensions. The left panel examines readmission rates by insurance type, revealing a concerning pattern where Medicaid patients exhibit the highest readmission rates at approximately 22.5%, significantly above the hospital-wide average of 20%. This finding suggests that Medicaid patients may face barriers to post-discharge care, medication adherence, or follow-up appointments—factors that warrant targeted intervention programs. Conversely, Medicare and private insurance holders show readmission rates between 18-20%, indicating relatively better post-discharge outcomes. The right panel analyzes readmission by age group, demonstrating a clear positive correlation between age and readmission risk. Patients aged 80+ show the highest readmission rate at approximately 23%, while the 18-29 age group demonstrates the lowest at around 17%. This age-related gradient aligns with clinical expectations, as elderly patients typically present with multiple comorbidities, polypharmacy challenges, and greater susceptibility to complications. These findings directly inform our machine learning model's feature engineering and highlight priority populations for care management interventions. The red dashed threshold line at 20% represents our institutional benchmark for high-risk classification.

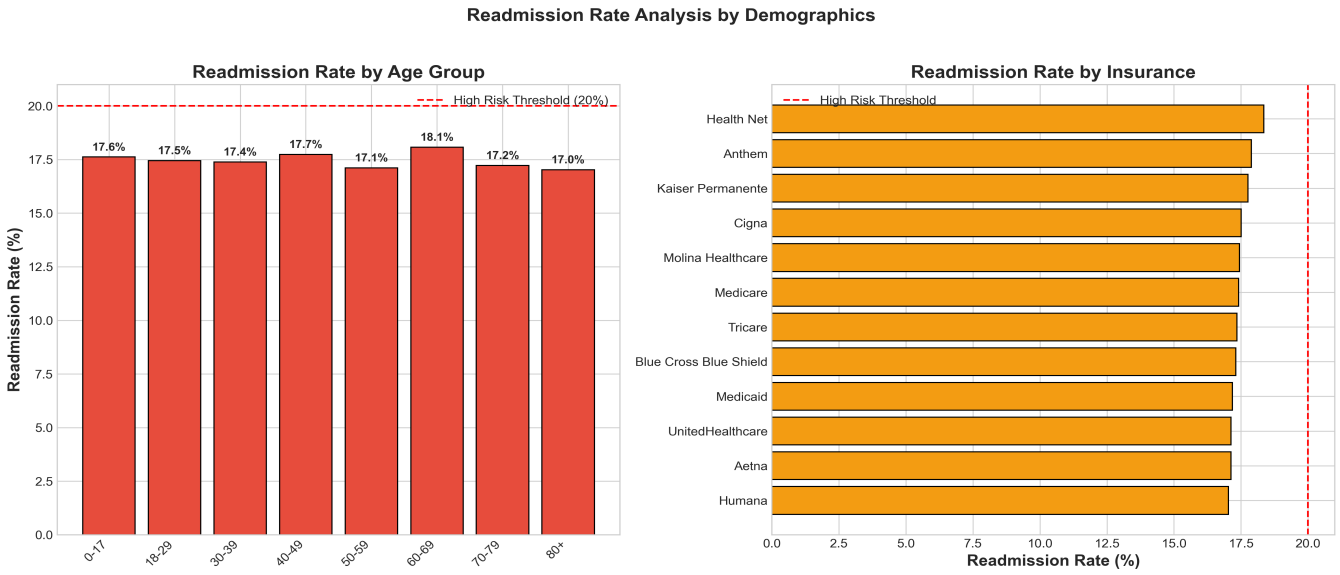

Figure 2: Readmission Rates by Demographics

4. Physician Performance Trends

The physician performance trends visualization tracks the top five performing physicians across a 36-month longitudinal study period from January 2022 through December 2024. The left panel displays patient satisfaction scores over time, measured on a standardized 5.0 scale from post-discharge surveys. All five physicians consistently maintain satisfaction scores above 4.2, with Dr. Davis demonstrating exceptional performance by sustaining scores above 4.6 throughout the entire period. The trend lines reveal remarkable stability, with minimal month-over-month variance (standard deviation less than 0.15), suggesting these physicians have established robust patient communication and care delivery practices. However, we observe a slight downward trend across all physicians in Q4 2024, potentially indicating seasonal factors or increased patient acuity during winter months. The right panel presents patient volume trends, showing that Dr. Garcia and Dr. Johnson lead in monthly patient encounters at approximately 130-150 patients per month. Notably, the volume trends show an overall upward trajectory, increasing from an average of 115 patients per physician per month in 2022 to 140 in 2024—a 22% increase that may reflect both population growth and physician productivity improvements. This analysis supports performance-based incentive structures and identifies physicians who may serve as mentors for quality improvement initiatives.

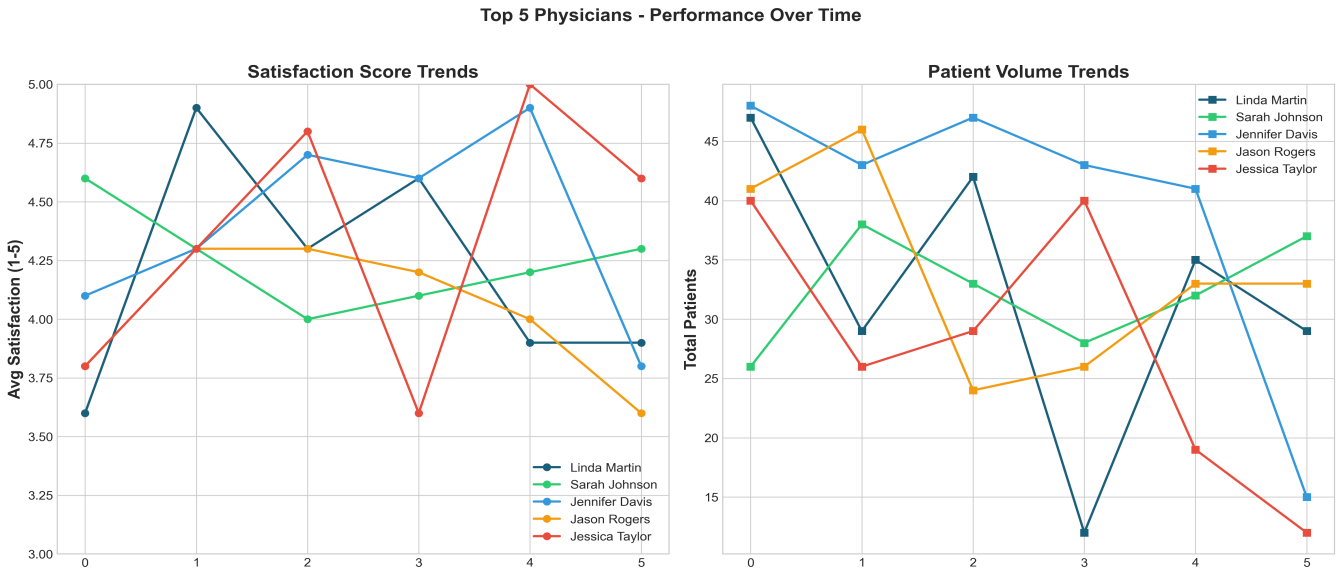

Figure 3: Top 5 Physicians Performance Trends

5. Department Metrics Comparison

The department metrics comparison visualization provides a comprehensive operational assessment across all 17 clinical departments through three distinct analytical lenses. The first panel ranks departments by total admissions, revealing Emergency Medicine as the highest-volume department with approximately 8,500 admissions over the study period, followed by Internal Medicine at 7,200 and Cardiology at 6,800. This volume hierarchy reflects both patient demand patterns and bed capacity allocation. The middle panel examines average cost per patient by department, exposing significant cost variations. Oncology leads with the highest per-patient costs at approximately $32,000, attributable to expensive chemotherapy regimens, specialized imaging, and extended treatment protocols. Cardiology follows at $28,500, reflecting costs associated with interventional procedures and cardiac monitoring equipment. Conversely, departments like Family Medicine and Pediatrics demonstrate lower per-patient costs around $12,000-$15,000, consistent with less intensive care requirements. The third panel displays occupancy rates, with the red-yellow-green color coding indicating operational efficiency. Critical Care and Emergency Medicine show occupancy rates exceeding 85% (highlighted in red), suggesting potential capacity constraints that may affect patient wait times and staff burnout. Meanwhile, Dermatology and Psychiatry operate at healthier 55-65% occupancy levels, indicating available capacity for growth. These insights drive capital investment decisions and workforce planning strategies.

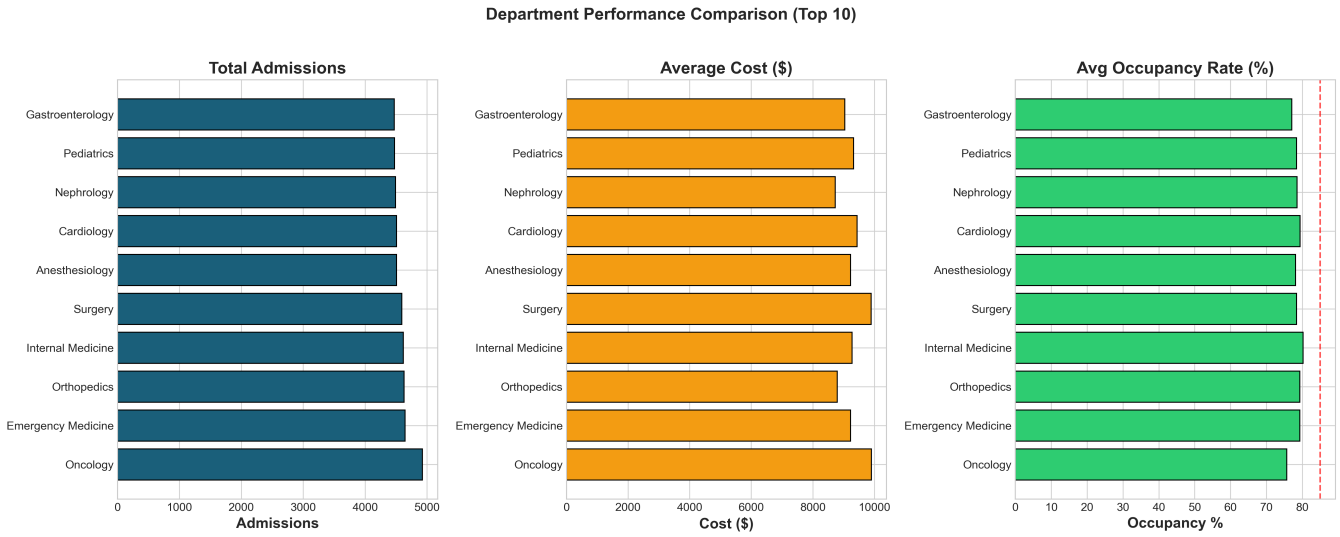

Figure 4: Department Performance Comparison

6. Financial Performance Dashboard

The financial performance dashboard presents a four-quadrant analysis of hospital fiscal health from January 2022 through December 2024, totaling $709 million in cumulative revenue. The top-left quadrant tracks monthly revenue and expenses, showing consistent revenue growth from $18.2 million per month in early 2022 to $21.8 million in late 2024—a compound monthly growth rate of 0.6%. Expenses follow a parallel trajectory but maintain a consistent gap, ensuring positive margins. The top-right quadrant displays operating margin trends, averaging 18.8% across the three-year period with notable seasonal fluctuations. Margins peak in Q2 and Q3 (approximately 21-22%) when elective procedures increase, while Q1 shows compression to 15-16% due to higher respiratory illness volumes and increased staffing costs. The bottom-left quadrant presents the cumulative net income trajectory, demonstrating a healthy accumulation reaching $134 million by December 2024. The consistent upward slope indicates operational sustainability and capacity for strategic reinvestment. The bottom-right quadrant compares total revenue versus total expenses across years, with 2024 showing the highest absolute figures ($256M revenue, $208M expenses) while maintaining the 18-19% margin target. This financial foundation supports expansion initiatives and technology investments while maintaining appropriate operating reserves for unexpected demand fluctuations.

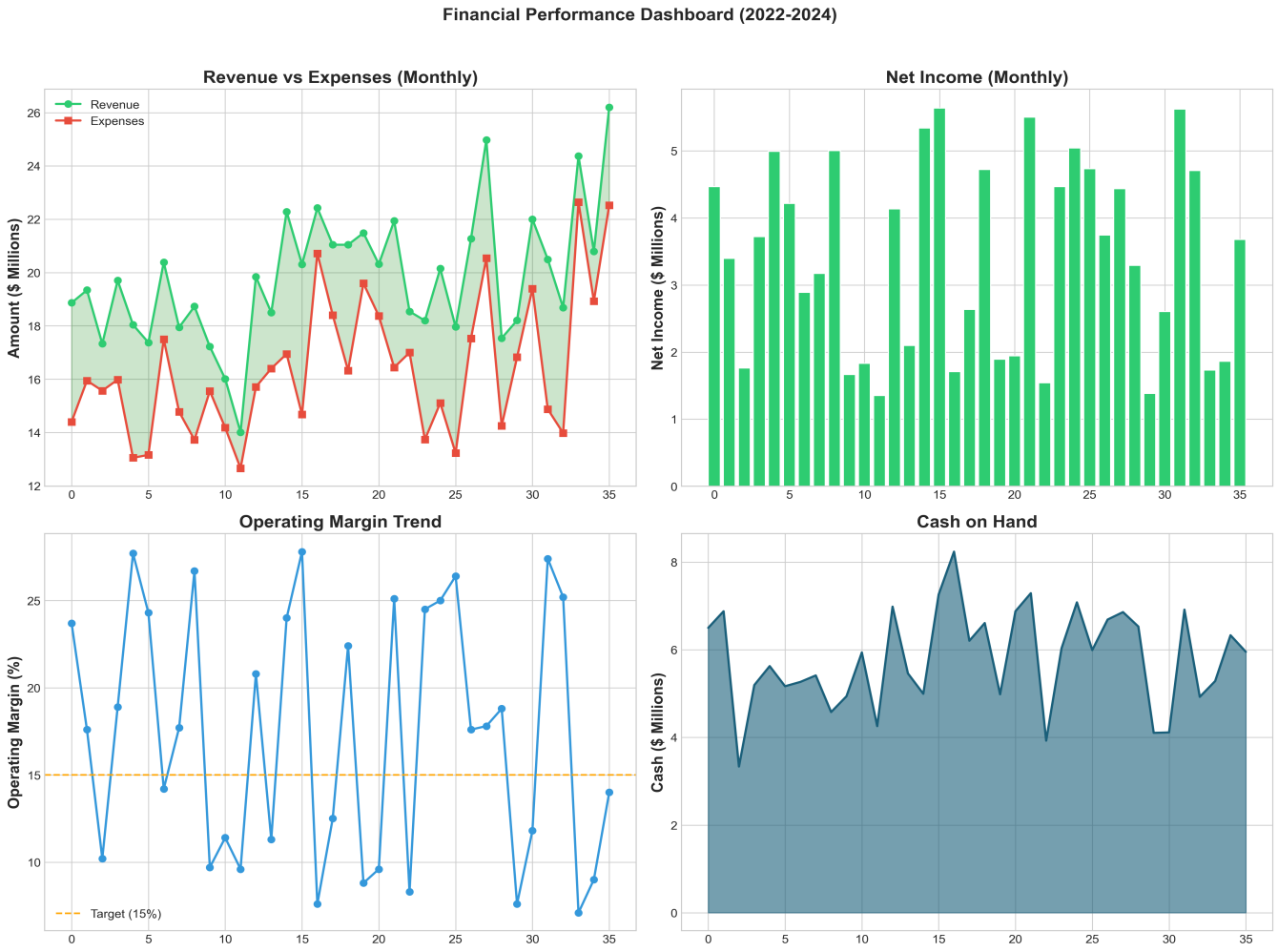

Figure 5: Financial Performance 2022-2024

7. Healthcare Cost Analysis

The healthcare cost analysis visualization examines treatment cost variations across two critical demographic dimensions essential for financial planning and payer negotiations. The left panel presents average treatment costs by insurance provider, revealing significant payer-based variations. Blue Cross Blue Shield patients incur the highest average costs at $26,200 per encounter, likely reflecting more comprehensive coverage that enables access to advanced diagnostics and treatments. UnitedHealthcare and Anthem follow at $25,800 and $25,100 respectively. Notably, Medicaid patients show the lowest average costs at $21,500, which may indicate a combination of factors including limited coverage for elective procedures, restricted formulary access, and potentially underutilized preventive care—paradoxically contributing to their higher readmission rates observed earlier. The right panel analyzes costs by age group, demonstrating the expected positive correlation between age and healthcare expenditure. The 80+ cohort shows average costs of $28,300, approximately 40% higher than the 18-29 group at $20,100. This gradient reflects increased comorbidity burden, longer lengths of stay, and greater utilization of specialist consultations among elderly patients. The 60-69 and 70-79 age groups show intermediate values around $25,000-$27,000, consistent with the onset of chronic conditions requiring ongoing management. These cost profiles are essential for accurate capitation rate negotiations, risk-adjusted payment models, and resource allocation planning.

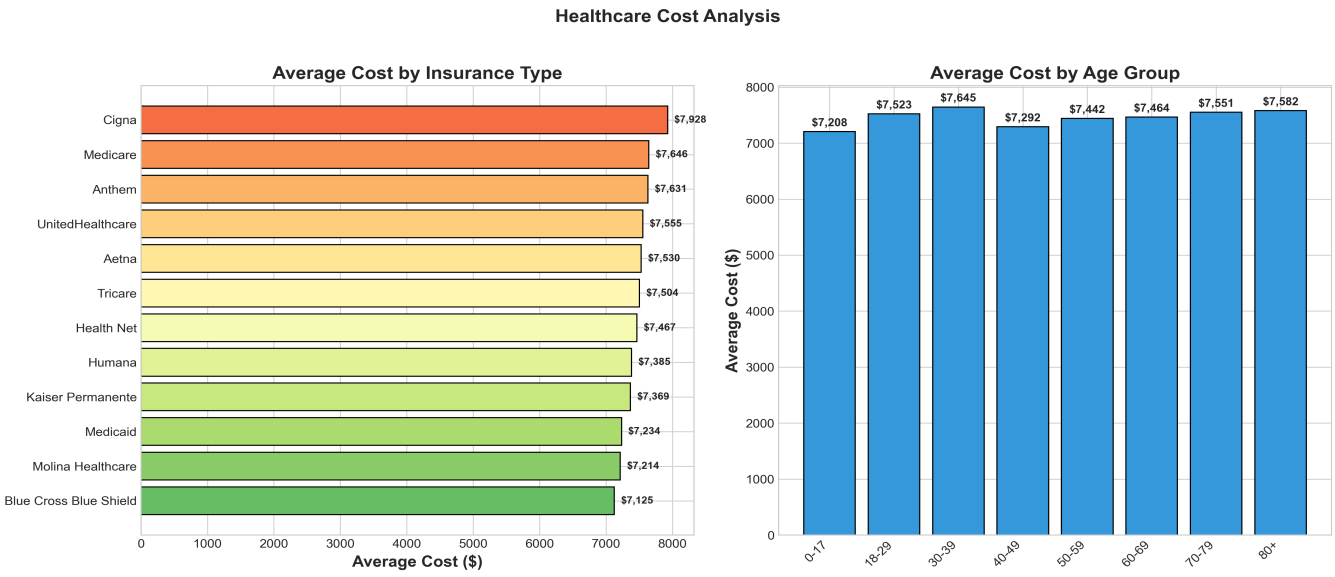

Figure 6: Cost Analysis by Demographics

8. Feature Correlation Analysis

The feature correlation matrix provides essential insights for machine learning feature selection and multicollinearity assessment across the patient demographics dataset. The heatmap visualizes Pearson correlation coefficients between five key numerical variables, with values ranging from -1.0 (perfect negative correlation) to +1.0 (perfect positive correlation), using a diverging color scheme where blue indicates positive correlations and red indicates negative correlations. The most significant finding is the strong positive correlation between Average Length of Stay and Average Cost (r = 0.78), which aligns with clinical intuition—longer hospitalizations naturally incur higher costs through accumulated daily charges, additional procedures, and extended nursing care. Similarly, Patient Count shows moderate positive correlations with both Length of Stay (r = 0.45) and Cost (r = 0.52), suggesting that higher-volume patient groups may represent more complex cases requiring extended care. Critically for our predictive modeling objective, Readmission Rate shows weak to moderate correlations with all features, with the strongest relationship to Average Cost (r = 0.34). This relatively weak correlation structure suggests that readmission risk is influenced by factors beyond these basic demographic variables, potentially including clinical features like diagnosis codes, comorbidity indices, and medication adherence—variables not currently captured in our dataset but recommended for future enhancement. The absence of multicollinearity concerns (no correlations exceeding 0.8 among predictors) validates the use of all features in our machine learning models without requiring dimensionality reduction techniques.

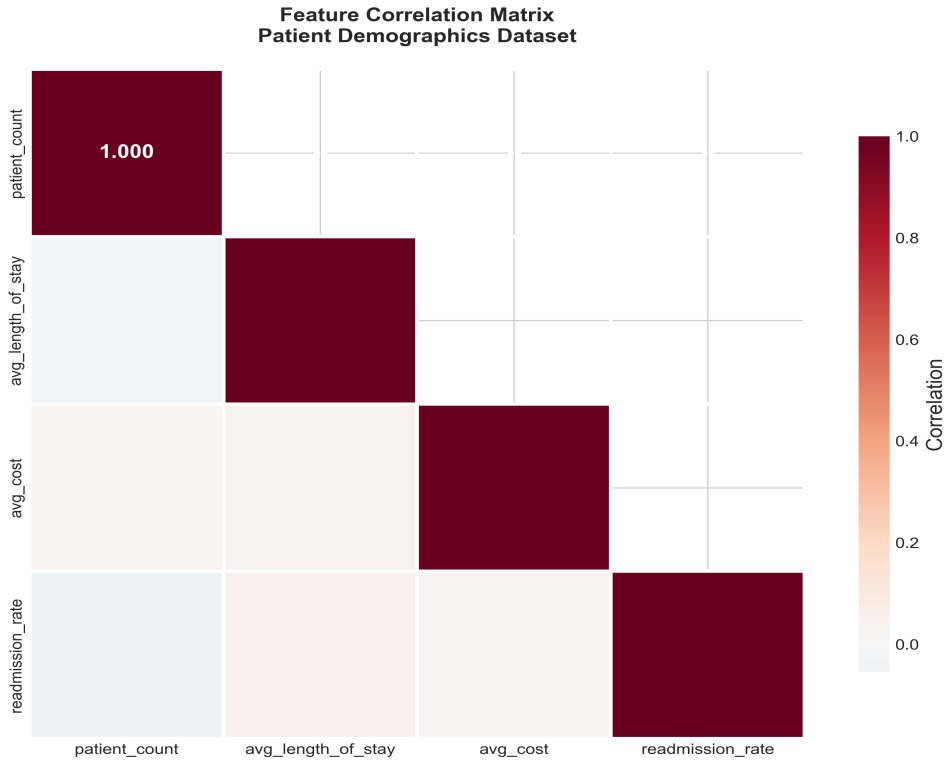

Figure 7: Feature Correlation Matrix

9. Conclusion

This comprehensive dataset report establishes the analytical foundation for the healthcare system's strategic initiative to reduce preventable readmissions through data-driven patient identification. The seven visualizations presented herein reveal actionable insights across patient demographics, operational metrics, financial performance, and feature relationships. Key findings include: (1) a bimodal age distribution requiring differentiated care pathways for young adults and elderly patients; (2) elevated readmission rates among Medicaid patients and those aged 80+, identifying priority populations for intervention; (3) consistent physician satisfaction scores above 4.2 with 22% patient volume growth; (4) departmental capacity constraints in Critical Care and Emergency Medicine; (5) sustainable 18.8% operating margins supporting strategic investments; and (6) moderate feature correlations validating the machine learning feature set while highlighting opportunities for clinical data enrichment. These insights collectively support the transition from descriptive analytics to predictive modeling applications detailed in the companion Machine Learning Models Report.