

Optimizing Hospital Resource Allocation Using Artificial Intelligence and Predictive Analytics

Phase 1: Data Understanding, Acquisition, and Feature Design

Dataset Overview

The analysis uses the CDC Weekly Hospital Respiratory Data (HRD) Metrics by Jurisdiction dataset. The uploaded file contains 20,770 weekly observations and 333 variables spanning 2020-08-08 through 2026-07-11. The file contains 67 geographic series, including states and territories, 10 HHS regions, and a national series.

Purpose and Suitability

The dataset was selected because it contains direct measures of inpatient and ICU occupancy, respiratory admissions, hospitalizations, and reporting coverage. These measures support one-week-ahead forecasting of hospital capacity pressure and can be connected to an optimization model for proactive bed and staffing allocation.

Data Quality Findings

No duplicate geography-week records were identified. Occupancy percentage variables provide substantially better continuity than many raw bed-count fields, so the initial forecasting design uses percentage occupancy measures as the primary operational targets. Historical count variables containing the value 999 were treated as unavailable in the analytical file, and percentage values outside the 0–100 range were excluded from modeling while retained in the audit documentation. Reporting coverage variables will remain available as data-quality controls.

Initial Feature Engineering

- Respiratory Admissions Total: combined COVID-19, influenza, and RSV weekly admissions.
- Respiratory Hospitalized Total: combined currently hospitalized respiratory patients.
- Operational Stress Index: weighted combination of inpatient occupancy, ICU occupancy, and respiratory admissions burden.
- Capacity Pressure: Low (<70%), Moderate (70–<80%), High (80–<90%), or Critical (≥90%) based on inpatient occupancy.
- Lag Features: one-week and two-week prior occupancy and admission values within each jurisdiction.
- Rolling Features: four-week historical averages using prior observations only.
- Forecast Targets: next-week inpatient occupancy percentage and next-week capacity-pressure category.

Modeling Panel Design

The clean modeling panel contains 17,360 state/territory-week observations. National and HHS regional rows are retained for descriptive benchmarking but excluded from the initial modeling panel to reduce

aggregation overlap. Each jurisdiction is ordered chronologically, and all lag and rolling features are calculated within jurisdiction to prevent cross-jurisdiction contamination.

Planned Forecasting Objective

The principal continuous target is next-week inpatient occupancy. A complementary classification target groups next-week capacity pressure into Low, Moderate, High, and Critical categories. The forecasting stage will compare a seasonal baseline, linear regression, Random Forest, XGBoost, and an appropriate time-series model. The optimization stage will translate the selected forecast into recommended beds, surge capacity, and staffing requirements under operational constraints.

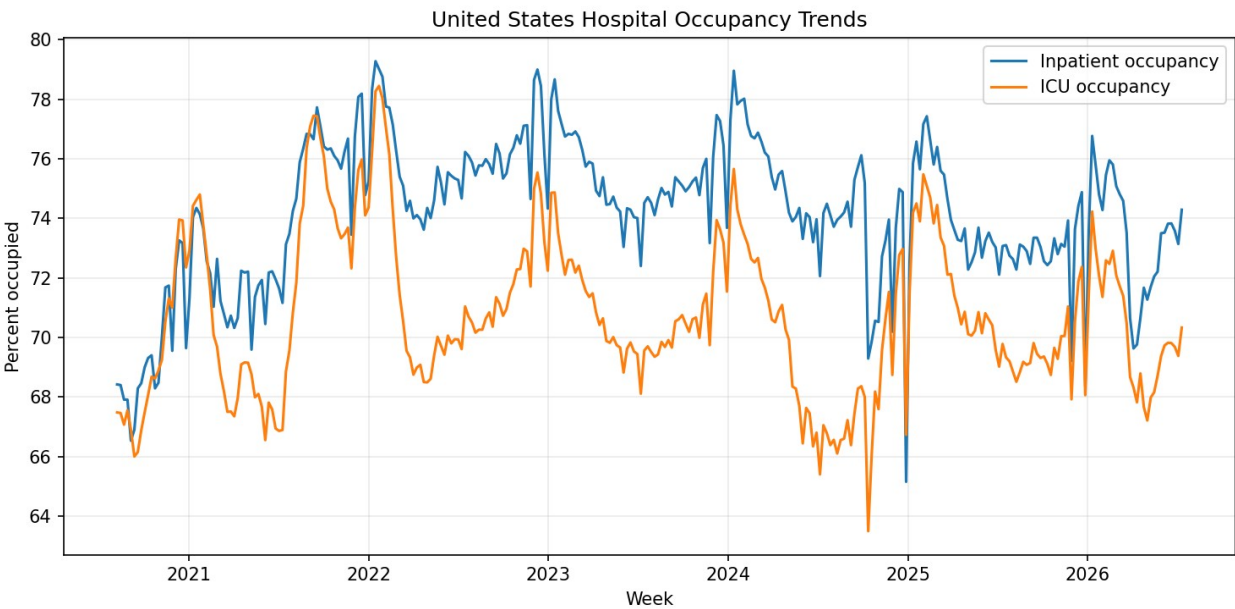


Figure 1. United States inpatient and ICU occupancy trends.

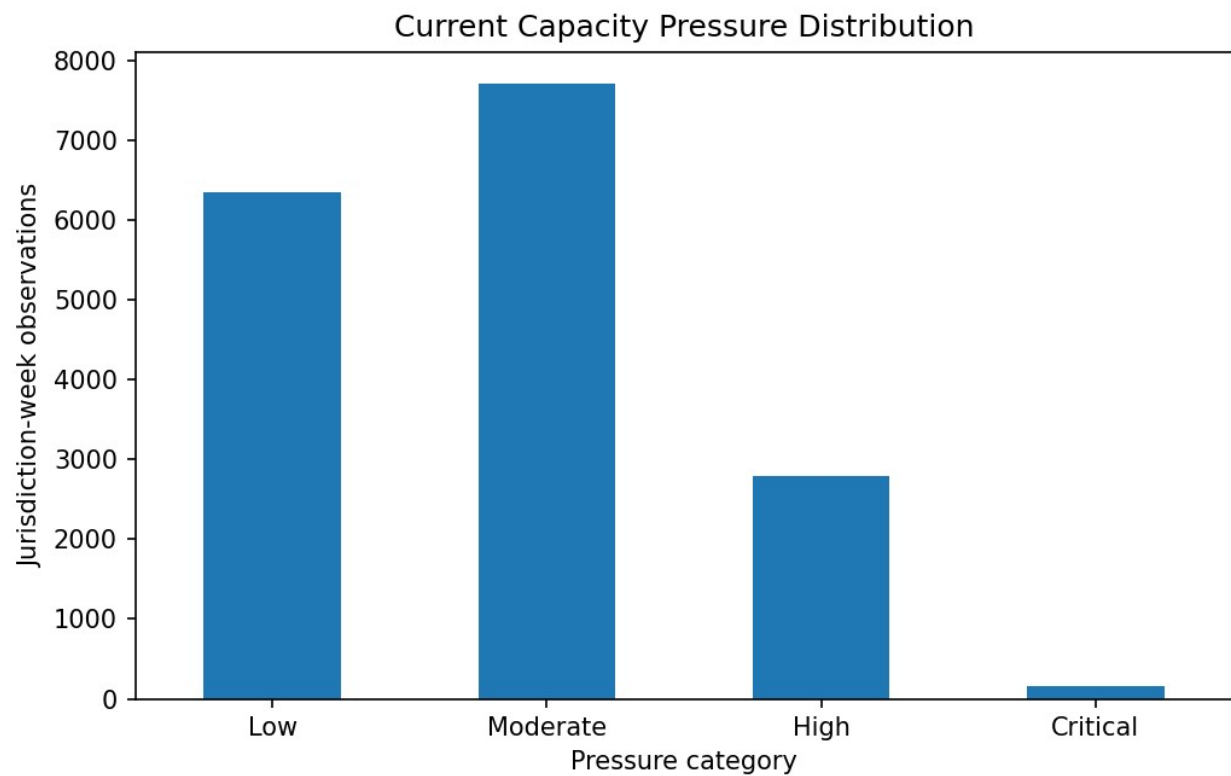


Figure 2. Distribution of jurisdiction-week capacity pressure.

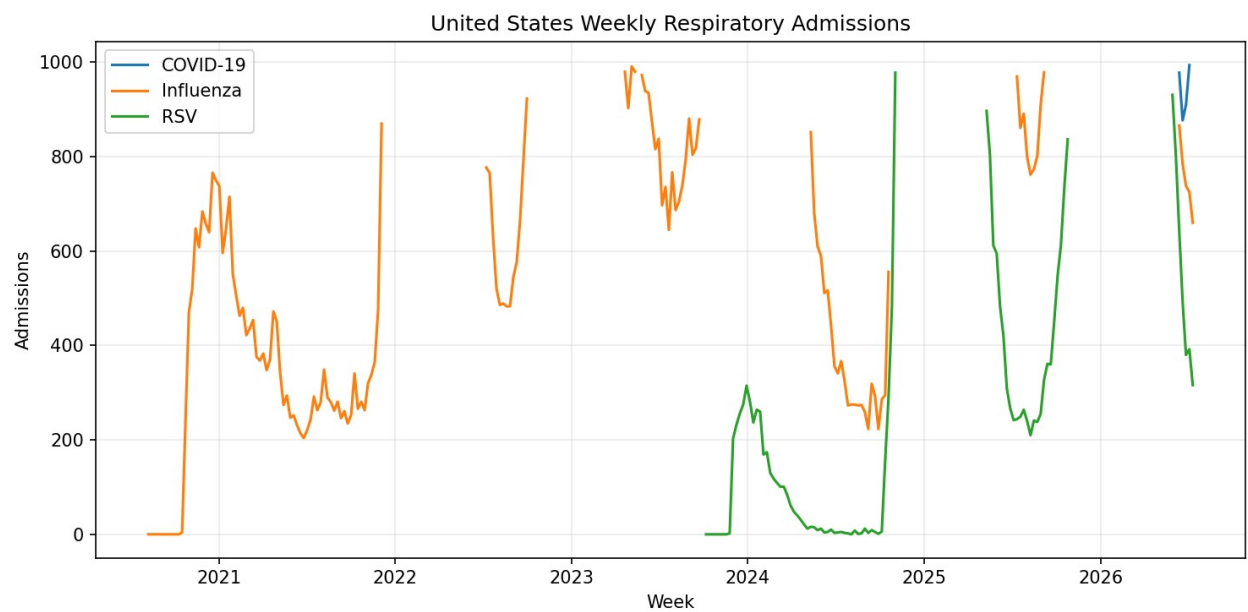


Figure 3. United States weekly respiratory admissions.

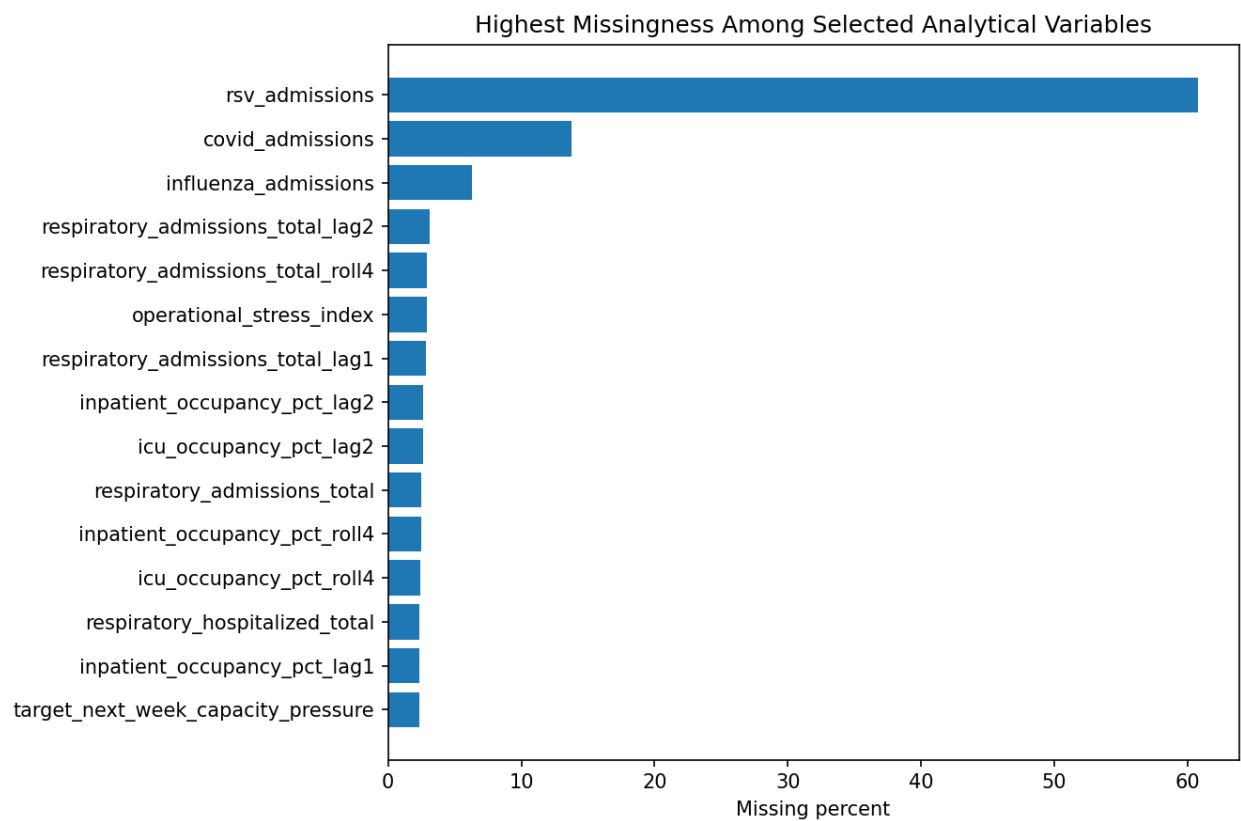


Figure 4. Missingness among selected analytical variables.

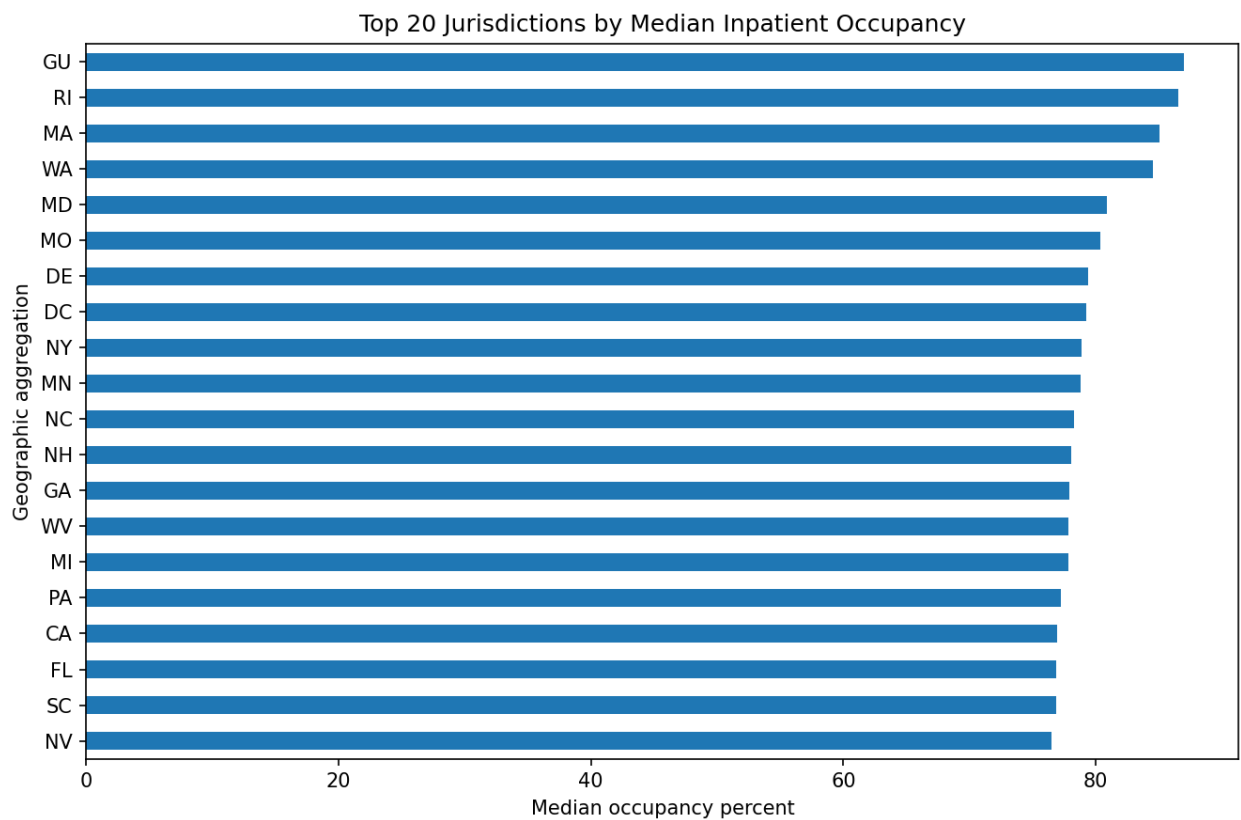


Figure 5. Jurisdictions with the highest median inpatient occupancy.

Phase 1 Conclusion

The dataset is suitable for forecasting and hospital resource-allocation research, but its reporting structure requires careful handling. The initial design prioritizes occupancy percentages, reporting-completeness controls, jurisdiction-specific lags, and transparent capacity-pressure thresholds. This Phase 1 package establishes the audited data foundation needed for exploratory analysis, forecasting, optimization, stress testing, and deployment planning.