Project: Hospital Readmission Analysis
Dataset Version: v1
Last Updated: January 2026
Prepared By: Michelle Obianwu

- Number of rows: 5001 rows
- Number of columns: 8 columns

Column Name	Data Type	Description	Example Value	Notes
Age	Categorical	Patient Age Group	69	
Gender	Categorical	Patient Gender	female	
Primary_Diagnosis	Categorical	Primary Diagnosis	Diabetes	
Num_Procedures	Numeric	Number of procedures during admission	1	
Days_in_Hospital	Numeric	Length of Stay (Days)	13	
Comorbidity_Score	Numeric	Clinical Risk Score	4	
Discharge_To	Categorical	Discharge Destination	Home Health Care	
Readmitted (Number)	Boolean (0/1)	Readmitted within 30 days (Yes/No)	0/1	

- Time range (if dates exist): No Dates Exist
- Source: Kaggle – Hospital Readmission Prediction Dataset

Business Problem

Hospital readmissions are a major quality and cost concern for healthcare organizations. This analysis evaluates patient demographics and clinical factors associated with hospital readmissions in order to identify high-risk populations and opportunities for improved care coordination.

Business Questions

1. Which patient groups have the highest hospital readmission rates?
2. Which diagnoses and procedure volumes are associated with increased readmission risk?
3. Which demographic groups represent the highest-risk populations for readmission?

Objectives

- Identify high-risk patient segments
- Evaluate clinical drivers of readmissions
- Support quality improvement initiatives

Missing Values Check:

All columns were reviewed using Excel filters.

No missing, blank, null, NA, or N/A values were found.

No rows were removed.

Data Cleaning & Quality Assurance

The dataset was reviewed and prepared for analysis using Microsoft Excel. A full data quality audit was performed prior to analysis.

Data Quality Checks

- All columns were reviewed for missing, null, and blank values using Excel filters.
- No missing or null values were found across the dataset.
- No duplicate records were identified.

Standardization

- Column headers were standardized for consistent naming conventions.
- Categorical fields were reviewed for inconsistent labels and spelling variations.
- No inconsistent category values were identified.

Validation

- Numeric fields (Number of Procedures, Days in Hospital, Comorbidity Score) were validated to ensure values were greater than or equal to zero.
- Readmission field was validated to contain only binary values (0 = Not Readmitted, 1 = Readmitted).

Final Dataset

The final cleaned dataset contains:

- 5,001 patient records
- 8 validated fields
- No missing values
- No invalid data

This dataset was approved for analysis and dashboard development in Tableau.

Readmission By Age Analysis

The age group that has the highest Readmission Rate is 29 years old at 29.73%

The age group that has the lowest readmission rate is 75 years old at 9.86%

Overall, readmission rates remain relatively consistent across age groups, suggesting that age alone is not a strong independent predictor of hospital readmission risk in this dataset.

Readmission By Diagnosis Analysis

The diagnosis with the highest hospital readmission rate is **Kidney Disease** at **19.90%**, while the lowest readmission rate is observed among patients with **Hypertension** at **17.79%**.

The top three diagnoses associated with the highest readmission risk are:

1. **Kidney Disease** — 19.90%
2. **Diabetes** — 19.47%
3. **COPD** — 18.53%

These findings indicate that chronic medical conditions are a significant driver of hospital readmissions, highlighting the importance of targeted care management programs for high-risk patient populations.

Readmissions by Number of Procedures Analysis

The highest hospital readmission rate is observed among patients who underwent 5 procedures, with a readmission rate of 21.01%. The lowest readmission rate occurs among patients with 0 procedures, at 16.99%.

Overall, readmission risk does not consistently increase with procedure volume. While patients with higher procedure counts might be expected to exhibit higher clinical complexity, the data shows that patients with 9 procedures have one of the lowest readmission rates.

These findings suggest that procedure volume alone is not a strong predictor of hospital readmission risk and that other clinical and demographic factors likely play a more significant role.

Readmissions by Length of Stay

The highest hospital readmission rate is observed among patients with a 7-day hospital stay, at 23.61%. The lowest readmission rate occurs among patients with a 13-day hospital stay, at 15.93%.

Overall, readmission risk does not consistently increase with length of stay. Patients with longer hospital stays do not demonstrate a significantly higher risk of readmission compared to those with shorter stays.

These findings suggest that length of stay alone is not a strong predictor of hospital readmissions, and that other clinical and demographic factors play a more meaningful role.

Executive Summary

Hospital readmissions are a major quality and cost concern for healthcare organizations. This analysis evaluated 5,001 hospital patient records to identify demographic and clinical factors associated with 30-day readmissions.

The analysis found that readmission rates remain relatively consistent across age groups and length of stay, indicating that these factors alone are not strong predictors of readmission risk. Chronic medical conditions, including kidney disease, diabetes, and COPD, were identified as the highest-risk diagnoses for hospital readmissions.

Procedure volume and length of stay were not found to be strong independent drivers of readmission risk, suggesting that readmissions are more closely linked to underlying clinical conditions rather than hospitalization intensity.

These findings support the need for targeted care management programs focused on patients with chronic diseases in order to reduce avoidable hospital readmissions and improve quality outcomes.

Key Performance Indicators (KPIs)

1. Readmission Rate

Definition:
Percentage of patients readmitted within 30 days of discharge.

Formula:
Total Readmissions ÷ Total Discharges

Business Use:
Used by hospital quality teams to monitor care effectiveness and CMS compliance.

2. High-Risk Patient Groups

Definition:
Patient populations with above-average readmission rates based on diagnosis, demographics, and clinical complexity.

Identified High-Risk Groups:

- Patients with kidney disease
- Patients with diabetes
- Patients with COPD

Business Use:
Used to prioritize care coordination, discharge planning, and follow-up interventions.

3. Cost Impact (Estimated)

Definition:
Estimated financial impact of hospital readmissions on operating costs.

Industry Benchmark:
Average hospital readmission cost: $15,000 per readmission

Estimated Impact Formula:
Number of Readmissions × Average Cost per Readmission

Business Use:
Used to justify investment in care management and readmission reduction programs.

Business Insights

- Readmission rates do not significantly vary by age or length of stay.
- Chronic conditions are the primary drivers of hospital readmissions.
- Procedure volume is not a reliable predictor of readmission risk.
- Targeted care management programs should focus on patients with chronic diseases to reduce avoidable readmissions.
- Reducing readmissions can significantly lower hospital operating costs and improve quality performance metrics.