**Diabetic Hospital Readmission Analysis**

**Team Members:** Himanshu Sharma, Tushar E Malankar, Sai Nithin Godi

# Introduction

Diabetic hospital readmissions strain healthcare systems, escalating costs and exposing gaps in care delivery. This project analyzes the UCI Diabetes 130-US Hospitals dataset, a publicly obtainable dataset, to uncover readmission patterns and deliver actionable insights. Using graph algorithms, logistic regression, and linear programming, we aim to:

* Identify high-risk care pathways.
* Predict readmission risks using clinical and demographic factors.
* Optimize follow-up care allocation.

These objectives address critical healthcare challenges, offering solutions to reduce readmissions, enhance patient outcomes, and improve hospital efficiency. The analysis provides evidence-based recommendations for providers, policymakers, and researchers. This report presents five key insights, supported by visualizations, detailing objectives, methods, findings, and implications for diabetic care.

# Dataset Description

The UCI Diabetes 130-US Hospitals dataset includes 101,766 diabetic patient encounter records from 130 US hospitals (1999–2008). Key features are:

* Admission Details: Source (e.g., Emergency Room), type (e.g., Emergency), discharge disposition (e.g., Home, Skilled Nursing Facility).
* Medication Changes: Diabetic medication adjustments (Change/No Change).
* Diagnoses: ICD9-coded, grouped (e.g., Circulatory, Respiratory).
* Readmitted: Target variable (<30 days, >30 days, No readmission).
* Demographics: Age (10-year intervals), race, gender.
* Clinical Metrics: Lab procedures, medications, hospital stay days.

Processed with data analysis tools, the dataset supports pathway analysis, prediction, and optimization. Missing values (e.g., race) were handled for robust analysis.

# Key Insights

The following insights obtained from each technique illuminate critical aspects of diabetic patient care, each structured with an objective, methodology, findings, and implications, supported by visualizations.

## **Technique 1: Logistic Regression for Readmission Prediction**

### **Objective**

To predict the likelihood of a diabetic patient being readmitted within 30 days of discharge and identify key risk factors for targeted interventions.

### **Key Insights**

* **Discharge types, admission sources, and diagnosis groupings are significant predictors of readmission risk**, with an ROC AUC of 0.687 indicating moderate predictive power.
* **Emergency admissions are strongly associated with higher readmission rates**, highlighting a critical entry point for intervention.

### **How Insights Were Derived**

A logistic regression model was trained on structured clinical and demographic data, including features like admission sources, discharge types, and diagnosis groupings. The model’s performance was evaluated using accuracy (64%), precision, recall, F1-score, and ROC AUC (0.687). A convergence warning suggested potential improvements through data scaling or increased iterations, but the model successfully identified key predictors. Analysis of feature coefficients revealed that emergency admissions, specific discharge dispositions, and certain diagnosis groups significantly influenced readmission likelihood.

### **Implications**

* **Clinical Prioritization**: Risk scores from the model can guide clinicians to prioritize high-risk patients, particularly those admitted through emergency pathways, for follow-up care.
* **Care Protocol Development**: The association of emergency admissions with readmissions suggests a need for standardized protocols to address this high-risk entry point.

## **Technique 2: Integer Linear Programming for Patient Triage**

### **Objective**

To select high-risk diabetic patients for follow-up interventions, optimizing readmission risk reduction within a $10,000 budget and operational constraints. However, the $10,000 budget constraint can be different from the actual budget.

### **Key Insights**

* **Optimal selection of 50 high-priority patients maximizes risk reduction** within the budget, focusing on those with elevated readmission risk.
* **Patients with emergency admissions or complex diagnoses are prioritized**, aligning with logistic regression findings.

### **How Insights Were Derived**

An Integer Linear Programming (ILP) model was formulated using risk scores from the logistic regression model as inputs. The ILP maximized readmission risk reduction by selecting patients for follow-up care while adhering to $10,000 budget and operational limits. The solution identified 50 patients, with analysis showing that those with emergency admissions or complex diagnosis profiles (e.g., circulatory or respiratory comorbidities) were consistently prioritized due to their higher risk scores.

### **Implications**

* **Targeted Resource Allocation**: The ILP ensures resources are directed to the highest-risk patients, potentially reducing readmissions cost-effectively.
* **Operational Flexibility**: The model can adapt to varying budgets or constraints, making it scalable for different hospital settings.

## **Technique 3: Linear Programming for Follow-Up Care Resource Allocation**

### **Objective**

To allocate 600 follow-up care hours across hospital departments to minimize costs while respecting departmental capacity constraints.

### **Key Insights**

* **Optimal allocation of 150 hours to Cardiology, 250 to Endocrinology, and 200 to Nephrology** achieves a total cost of $61,000, balancing cost and capacity.
* **Endocrinology receives the largest share**, reflecting its critical role in diabetes management.

### **How Insights Were Derived**

A Linear Programming (LP) model was developed to allocate 600 follow-up care hours across Cardiology, Endocrinology, and Nephrology. The model minimized total care delivery costs while ensuring departmental capacity limits were met. The solution allocated 250 hours to Endocrinology, 200 to Nephrology, and 150 to Cardiology, reflecting the higher relevance of Endocrinology and Nephrology to diabetic care. The total cost of $61,000 was derived by optimizing cost coefficients against capacity constraints.

### **Implications**

* **Cost-Effective Care Delivery**: The LP model ensures efficient use of follow-up care hours, maximizing impact within budget constraints.
* **Departmental Resource Planning**: Hospitals can adjust staffing or training in Endocrinology and Nephrology to meet the needs of diabetic patients.

## **Technique 4: Graph Algorithms for Diagnosis Co-Occurrence Network Analysis**

### **Objective**

To identify common comorbidity clusters among diabetic patients to support the development of bundled care pathways.

### **Key Insights**

* **Circulatory and Respiratory conditions are the most central diagnosis groups**, frequently co-occurring with Diabetes.
* **Strong links between Diabetes, Circulatory, and Respiratory diagnoses** indicate high-risk comorbidity clusters.

### **How Insights Were Derived**

An undirected co-occurrence network was constructed using graph algorithms, with nodes representing diagnosis groups and edges indicating co-occurrence in patient records. Network analysis, implemented via NetworkX, calculated centrality measures to identify the most connected diagnosis groups. Circulatory and Respiratory conditions emerged as highly central, with strong edges to Diabetes, indicating frequent co-occurrence in patient profiles.

### **Implications**

* **Bundled Care Pathways**: Developing integrated care protocols for Diabetes, Circulatory, and Respiratory conditions can improve outcomes for high-risk patients.
* **Resource Allocation**: Hospitals can prioritize resources for managing these comorbidity clusters, potentially reducing readmission rates.
* **Preventive Measures**: Insights into comorbidity patterns support proactive screening and early intervention to address co-occurring conditions.

**Visualization**

**A diagram of a network

AI-generated content may be incorrect. A diagram of a network

AI-generated content may be incorrect.**

## **Technique 5: Graph Algorithms for Extended Patient Pathway Network Analysis**

### **Objective**

To map patient care pathways from admission to discharge, identifying common patterns to inform process standardization and reduce readmissions.

### **Key Insights**

* **Emergency Room Referral → Emergency Admission → Med Change No → Discharged to Home** is the most common pathway (15,480 patients), followed by the same pathway with medication changes (13,245 patients).
* **Variations in medication change status** across pathways suggest inconsistent treatment practices that may impact readmission risk.

### **How Insights Were Derived**

A directed patient pathway network was built using graph algorithms, tracing the sequence of **Admission Source → Admission Type → Medication Change → Discharge**. Patient records were processed to extract transitions, and a directed graph was constructed using NetworkX. The top five pathways were identified by counting complete pathways, with the most frequent involving emergency room referrals and discharges to home. Analysis of medication change status revealed variations within these pathways.

### **Implications**

* **Care Standardization**: Standardizing treatment protocols, particularly around medication changes for emergency admissions, could reduce readmission risk.
* **Intervention Focus**: Targeting interventions at emergency room referrals can address a significant portion of the patient population.
* **Discharge Planning**: Insights into discharge patterns support the development of robust discharge protocols to ensure continuity of care and reduce readmissions.

**Visualization**

A diagram of a network with Camino de Santiago in the background

AI-generated content may be incorrect.

# Conclusion

This analysis of the UCI Diabetes 130-US Hospitals dataset provides a comprehensive framework for reducing diabetic readmissions. Graph algorithms, logistic regression, and linear programming, supported by visualizations, revealed high-risk pathways, comorbidity patterns, risk predictors, and optimal resource strategies. These insights enable targeted interventions— emergency department protocols, refined medication management, integrated care pathways, and enhanced follow-up—while optimized resource allocation improves efficiency. The findings offer scalable solutions for clinical practices, cost reduction, and policy development. Future work could explore clustering to segment patient groups or decision trees to refine risk models, further advancing data-driven healthcare solutions with broad impact.