QMH Fest Data Analysis Competition 2025: Hospital Readmissions Data

Title:
Predicting and Reducing Hospital Readmissions Using Patient-Level Data

Background & Objective:

Hospital readmissions are a major concern for healthcare systems worldwide. High readmission rates often reflect gaps in the quality of care and can result in financial penalties under many healthcare insurance policies. Reducing unnecessary readmissions not only improves patient outcomes but also significantly lowers healthcare costs.

In this challenge, you are tasked with acting as a data analyst for a national health policy think tank. Your mission is to analyze patterns of hospital readmissions and design an intervention strategy informed by statistical modeling.

You are provided with a real-world hospital dataset containing patient demographics, medical history, visit details, and treatment information. Your goal is to build a data-driven approach that identifies key factors associated with hospital readmission and to propose targeted strategies to reduce preventable readmissions.

Given Dataset:

- A dataset (diabetic_data_QMH_Club_Fest_2025) with over 100,000 patient records
- Variables include:
  - Patient demographics (age, gender, race)
  - Admission and discharge details
  - Diagnoses and prior visit counts
  - Medications and lab test results
  - Readmission status (readmitted: "NO", ">30", "<30")

Tasks:

Participants are expected to complete the following:

1. ***Formulate a Research Question / Hypotheses***

2. ***Data Preparation***

3. ***Exploratory Data Analysis (EDA)***

4. ***Predictive Modeling***

5. ***Model Evaluation***

6. ***Interpretation***

7. ***Recommendation & Strategy***

Final Deliverables:

Participants must submit:

1. A concise report *(maximum 15 pages; font size 12, line spacing 1.5)* including:

   o Research question
   o Summary of methods
   o Key results and model performance
   o Interpretation of findings
   o Recommendations for interventions or policy

2. Code/scripts used
   R, Python, SAS, or SPSS code/scripts must be included for reproducibility.

Time Allotted:

Final submission deadline: 24 October 2025

Evaluation Criteria:

Criterion	Weight
Research question clarity	5%
Data cleaning & preparation	10%
Description and visualizations	20%
Model design and justification	30%
Interpretation of results	20%
Intervention strategy	5%
Presentation/report quality	10%

Variable Description:

Table 1: Variable description

Variable	Name	Type	Description	Missing(%)
V1	encounter_code *single visit*	Numeric	Unique identifier of an encounter	0%
V2	patient_code	Numeric	Unique identifier of a patient	0%
V3	ethnic_group	Nominal	Values: Caucasian, Asian, AfricanAmerican, Hispanic, and other	2%
V4	sex_identity	Nominal	Values: male, female, and unknown/invalid	0%
V5	age_band	Nominal	Grouped in 10-year intervals: [0-10), [10-20), ..., [90-100)	0%
V6	body_weight	Numeric	Weight in pounds	97%
V7	adm_type_code	Nominal	9 distinct values (emergency, urgent, elective, etc.)	0%
V8	discharge_type	Nominal	29 values (e.g., discharged to home, expired, etc.)	0%
V9	adm_source_code	Nominal	21 values (e.g., physician referral, ER, transfer from hospital)	0%
V10	hospital_days	Numeric	Number of days between admission and discharge	0%
V11	insurance_code	Nominal	23 values (e.g., Medicare, Medicaid, self-pay)	52%
V12	provider_specialty	Nominal	84 values (e.g., cardiology, internal medicine, etc.)	53%
V13	lab_test_count	Numeric	Number of lab tests performed	0%
V14	procedure_count	Numeric	Number of procedures (excluding labs)	0%
V15	medication_count	Numeric	Number of distinct medications administered	0%
V16	outpatient_visits	Numeric	Number of outpatient visits in the past year	0%
V17	emergency_visits	Numeric	Number of emergency visits in the past year	0%
V18	inpatient_visits	Numeric	Number of inpatient visits in the past year	0%
V19	diagnosis_primary	Nominal	Primary diagnosis (ICD-9 code, first 3 digits)	0%
V20	diagnosis_secondary	Nominal	Secondary diagnosis	0%
V21	diagnosis_tertiary	Nominal	Additional secondary diagnosis	1%

Variable	Name	Type	Description	Missing(%)
V22	diagnosis_total	Numeric	Number of diagnoses entered into the system	0%
V23	glucose_test_result	Nominal	Glucose serum result or "None"	0%
V24	A1C_result	Nominal	A1C test result: ">8", ">7", "Normal", "None"	0%
V25–V47	medication_columns	Nominal	For 24 drugs, value = "Up", "Down", "Steady", "No"	0%
V48	med_change_status	Nominal	Indicates if medication was changed: "Change", "No change"	0%
V49	diabetic_med_given	Nominal	Indicates if diabetes medication was prescribed: Yes/No	0%
V50	readmission_status	Nominal	Readmission: "<30", ">30", "NO"	0%

ID mapping

adm_type_code(V7)	Description
1	Emergency
2	Urgent
3	Elective
4	Newborn
5	Not Available
6	NULL
7	Trauma Center
8	Not Mapped

discharge_type (V8)	Description
1	Discharged to home
2	Discharged/transferred to another short term hospital
3	Discharged/transferred to SNF
4	Discharged/transferred to ICF
5	Discharged/transferred to another type of inpatient care institution
6	Discharged/transferred to home with home health service
7	Left AMA
8	Discharged/transferred to home under care of Home IV provider
9	Admitted as an inpatient to this hospital
10	Neonate discharged to another hospital for neonatal aftercare
11	Expired
12	Still patient or expected to return for outpatient services
13	Hospice / home
14	Hospice / medical facility
15	Discharged/transferred within this institution to Medicare approved swing bed
16	Discharged/transferred/referred another institution for outpatient services
17	Discharged/transferred/referred to this institution for outpatient services
18	NULL
19	Expired at home. Medicaid only, hospice.
20	Expired in a medical facility. Medicaid only, hospice.
21	Expired, place unknown. Medicaid only, hospice.
22	Discharged/transferred to another rehab fac including rehab units of a hospital .
23	Discharged/transferred to a long term care hospital.
24	Discharged/transferred to a nursing facility certified under Medicaid but not certified under Medicare.
25	Not Mapped
26	Unknown/Invalid
30	Discharged/transferred to another Type of Health Care Institution not Defined Elsewhere
27	Discharged/transferred to a federal health care facility.
28	Discharged/transferred/referred to a psychiatric hospital of psychiatric distinct part unit of a hospital
29	Discharged/transferred to a Critical Access Hospital (CAH).

admission_soure_code (V9)   Description

1	Physician Referral
2	Clinic Referral
3	HMO Referral
4	Transfer from a hospital
5	Transfer from a Skilled Nursing Facility (SNF)
6	Transfer from another health care facility
7	Emergency Room
8	Court/Law Enforcement
9	Not Available
10	Transfer from critial access hospital
11	Normal Delivery
12	Premature Delivery
13	Sick Baby
14	Extramural Birth
15	Not Available
17	NULL
18	Transfer From Another Home Health Agency
19	Readmission to Same Home Health Agency
20	Not Mapped
21	Unknown/Invalid
22	Transfer from hospital inpt/same fac reslt in a sep claim
23	Born inside this hospital
24	Born outside this hospital
25	Transfer from Ambulatory Surgery Center
26	Transfer from Hospice

Table 2: International Classification of Diseases (ICD-9) Diagnosis Group Code Description for V19-V21

Group Name	ICD-9 Code Range(s)	Description
Circulatory	390–459, 785	Diseases of the circulatory system
Respiratory	460–519, 786	Diseases of the respiratory system
Digestive	520–579, 787	Diseases of the digestive system
Diabetes	250.xx	Diabetes mellitus
Injury	800–999	Injury and poisoning
Musculoskeletal	710–739	Diseases of the musculoskeletal system and connective tissue
Genitourinary	580–629, 788	Diseases of the genitourinary system
Neoplasms	140–239	Neoplasms
Other (Symptoms)	780, 781, 784, 790–799	Other symptoms, signs, and ill-defined conditions
Endocrine (Excl. DM)	240–279 (excluding 250)	Endocrine, nutritional and metabolic diseases and immunity disorders (non-DM)
Skin/Subcutaneous	680–709, 782	Diseases of the skin and subcutaneous tissue
Infectious	001–139	Infectious and parasitic diseases
Mental	290–319	Mental disorders
External Causes	E–V codes	External causes of injury and supplemental classification
Blood Disorders	280–289	Diseases of the blood and blood-forming organs
Nervous System	320–359	Diseases of the nervous system
Pregnancy/Childbirth	630–679	Complications of pregnancy, childbirth, and the puerperium
Sense Organs	360–389	Diseases of the sense organs
Congenital Anomalies	740–759	Congenital anomalies

Notes:

- "250.xx" represents all subcodes of diabetes in ICD-9 (e.g., 250.00, 250.01...).
- The "Other" group includes general symptoms and ill-defined conditions not falling into major disease categories.
- External cause codes (E–V) are used to classify injury mechanisms and other external health influences.