Data-Driven Prediction of Hospital Readmissions

This project aims to predict hospital readmissions using patient data. By applying machine learning techniques, we identify patterns that help determine which patients are at risk of being readmitted, enabling better healthcare decisions and early interventions.

factors that lead to the high readmission rate of diabetic patients within 30 days post discharge and correspondingly to predict the high-risk diabetic-patients who are most likely to get readmitted within 30 days so that the quality of care can be improved along with improved patient's experience, health of the population and reduce costs by lowering readmission rates. Also, to identify the medicines that are the most effective in treating diabetes.

Step 1: data acquisition

click on the image to visit the dataset (CTRL click):

Step2: Data Interpretation:

Admission severity id (ranking 1-9 as shown below)

1.	Emergency	Immediate and critical care is required; life-threatening.
2.	Trauma Centre:	Urgent specialized care for severe injuries or trauma.
3.	Urgent:	Prompt medical attention needed but not immediately life-threatening.
4.	Newborn:	Priority care for newborns, typically time-sensitive.
5.	Transfer from Another Facility:	Urgent or necessary continuation of care from another institution.
6.	Outpatient Converted to Inpatient:	Admission indicates worsening of the condition requiring inpatient care.
7.	Observation:	Monitoring conditions that may not yet require full inpatient care.
8.	Elective:	Planned procedures; urgency is generally lower.
9.	Other:	Likely to include non-urgent or unspecified cases.

Interpretations for Discharge Disposition IDs (1–29):

1. Home-Based Dispositions**:** Discharged to home/self-care, Discharged to home with home healthcare
2. Discharged to home under custodial care
3. Transfer to Other Facilities:
4. Transferred to short-term rehabilitation
5. Transferred to long-term care facility
6. Transferred to another hospital
7. Transferred to a psychiatric facility
8. Transferred to intermediate care facility
9. Transferred to inpatient hospice
14–29. May include additional combinations

Interpreting diag_1(main reason for hospitalization, diag_2, and diag_3 (additional causes may worse the condition) (ICD-9 Codes)

ICD-9 Code Ranges and Their Meaning

This is the pdf link for detailed explanation of ICD-9 codes

ICD-9 Range	Condition Category	Example Code	Meaning
001–139	Infectious & Parasitic Diseases	038.9	Sepsis
140–239	Neoplasms (Tumors/Cancers)	174.9	Breast Cancer
240–279	Endocrine, Nutritional, & Metabolic Diseases	250.00	Diabetes Mellitus
280–289	Blood Disorders	285.9	Anemia
290–319	Mental Disorders	311	Depression
320–389	Nervous System Disorders	345.9	Epilepsy
390–459	Circulatory System Disorders	401.9	Hypertension
460–519	Respiratory System Disorders	486	Pneumonia
520–579	Digestive System Disorders	530.81	GERD (Acid Reflux)
580–629	Genitourinary System Disorders	585.9	Chronic Kidney Disease
630–679	Pregnancy-Related Conditions	644.21	Early Labor
680–709	Skin & Subcutaneous Conditions	682.9	Skin Infection
710–739	Musculoskeletal Disorders	715.9	Osteoarthritis
740–759	Congenital Anomalies	759.9	Congenital Malformation
760–779	Perinatal Conditions	765.1	Preterm Infant
780–799	Symptoms, Signs & Ill-Defined Conditions	786.5	Chest Pain
800–999	Injury & Poisoning	820.8	Hip Fracture
V01–V89	Factors Influencing Health	V58.67	Long-term Insulin Use
E800–E999	External Causes of Injury	E885	Fall Injury

Oral Diabetes Medications Interpretation (brief)

These drugs help control **blood sugar levels** in diabetes patients:

Medicine	Type	Function
Metformin	Biguanide	Reduces liver glucose production, increases insulin sensitivity.
Repaglinide	Meglitinide	Stimulates insulin release from the pancreas.
Nateglinide	Meglitinide	Similar to repaglinide, but shorter-acting.
Chlorpropamide	Sulfonylurea	Stimulates insulin secretion from the pancreas. (Older drug, rarely used now.)
Glimepiride	Sulfonylurea	Increases insulin secretion, longer-lasting effect.
Acetohexamide	Sulfonylurea	Rarely used now, similar to chlorpropamide.
Glipizide	Sulfonylurea	Short-acting insulin secretion stimulator.
Glyburide	Sulfonylurea	Similar to glipizide but longer-acting.
Tolbutamide	Sulfonylurea	Older, rarely used now.
Pioglitazone	Thiazolidinedione (TZD)	Improves insulin sensitivity in muscle and fat cells.
Rosiglitazone	Thiazolidinedione (TZD)	Similar to pioglitazone but has cardiovascular risk concerns.
Acarbose	Alpha-glucosidase inhibitor	Delays carbohydrate digestion, reducing blood sugar spikes.
Miglitol	Alpha-glucosidase inhibitor	Works like acarbose but with slightly different metabolism.
Troglitazone	Thiazolidinedione (TZD)	Withdrawn due to liver toxicity.
Tolazamide	Sulfonylurea	Another older insulin-secreting drug.

Combination Medications (mix of 2 or more)

These drugs combine different **mechanisms** to control blood sugar:

Medicine	Components
Glyburide-Metformin	Sulfonylurea + Metformin
Glipizide-Metformin	Sulfonylurea + Metformin
Glimepiride-Pioglitazone	Sulfonylurea + TZD
Metformin-Rosiglitazone	Biguanide + TZD
Metformin-Pioglitazone	Biguanide + TZD

Some of the Medical terminologies

1. diag_1: Primary diagnosis code indicating the main reason for the patient's hospital admission.
2. diag_2: Secondary diagnosis code representing an additional but relevant medical condition.
3. diag_3: Tertiary diagnosis code for another co-existing condition noted during admission.
4. Maximum_glucose_serum : Represents the **highest blood glucose level** measured during the hospital stay, recorded as a categorical value. ***'None'*** – (No test was performed.) ***'Norm'*** – (Glucose levels were normal.) '>***200'*** – (Glucose levels were greater than 200 mg/dL.) '>***300'*** – (Glucose levels were greater than 300 mg/dL.)
5. The **A1C** test measures a patient's average blood glucose level over the past 2–3 months, and is commonly used to diagnose and monitor diabetes. Test for glycated hemoglobulin means how much heme is glycated --> In the dataset (**A1Cresult** column):'None' – No A1C test was performed. '**Norm**' – A1C level was within the normal range. '>**7**' – A1C result was greater than 7%. '>**8**' – A1C result was greater than 8%.

Data description form the original source

Feature name	Type	Description and values	% missing
Encounter ID	Numeric	Unique identifier of an encounter	0%
Patient number	Numeric	Unique identifier of a patient	0%
Race	Nominal	Values: Caucasian, Asian, African American, Hispanic, and other	2%
Gender	Nominal	Values: male, female, and unknown/invalid	0%
Age	Nominal	Grouped in 10-year intervals: [0, 10), [10, 20), . . ., [90, 100)	0%
Weight	Numeric	Weight in pounds.	97%
Admission type	Nominal	Integer identifier corresponding to 9 distinct values, for example, emergency, urgent, elective, newborn, and not available	0%
Discharge disposition	Nominal	Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available	0%
Admission source	Nominal	Integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital	0%
Time in hospital	Numeric	Integer number of days between admission and discharge	0%
Payer code	Nominal	Integer identifier corresponding to 23 distinct values, for example, Blue Cross\Blue Shield, Medicare, and self-pay	52%
Medical specialty	Nominal	Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, cardiology, internal medicine, family\general practice, and surgeon	53%
Number of lab procedures	Numeric	Number of lab tests performed during the encounter	0%
Number of procedures	Numeric	Number of procedures (other than lab tests) performed during the encounter	0%
Number of medications	Numeric	Number of distinct generic names administered during the encounter	0%
Number of outpatient visits	Numeric	Number of outpatient visits of the patient in the year preceding the encounter	0%
Number of emergency visits	Numeric	Number of emergency visits of the patient in the year preceding the encounter	0%
Number of inpatient visits	Numeric	Number of inpatient visits of the patient in the year preceding the encounter	0%
Diagnosis 1	Nominal	The primary diagnosis (coded as first three digits of ICD9); 848 distinct values	0%
Diagnosis 2	Nominal	Secondary diagnosis (coded as first three digits of ICD9); 923 distinct values	0%
Diagnosis 3	Nominal	Additional secondary diagnosis (coded as first three digits of ICD9); 954 distinct values	1%
Number of diagnoses	Numeric	Number of diagnoses entered to the system	0%
Glucose serum test result	Nominal	Indicates the range of the result or if the test was not taken. Values: “>200,” “>300,” “normal,” and “none” if not measured	0%
A1c test result	Nominal	Indicates the range of the result or if the test was not taken. Values: “>8” if the result was greater than 8%, “>7” if the result was greater than 7% but less than 8%, “normal” if the result was less than 7%, and “none” if not measured.	0%
Change of medications	Nominal	Indicates if there was a change in diabetic medications (either dosage or generic name). Values: “change” and “no change”	0%
Diabetes medications	Nominal	Indicates if there was any diabetic medication prescribed. Values: “yes” and “no”	0%
24 features for medications	Nominal	For the generic names: metformin, repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, sitagliptin, insulin, glyburide-metformin, glipizide-metformin, glimepiride-pioglitazone, metformin-rosiglitazone, and metformin-pioglitazone, the feature indicates whether the drug was prescribed or there was a change in the dosage. Values: “up” if the dosage was increased during the encounter, “down” if the dosage was decreased, “steady” if the dosage did not change, and “no” if the drug was not prescribed	0%
Readmitted	Nominal	Days to inpatient readmission. Values: “<30” if the patient was readmitted in less than 30 days, “>30” if the patient was readmitted in more than 30 days, and “No” for no record of readmission.	0%

Step3: preprocessing

a. Deleting irrelevant cols (eg. Encounter_id, etc.)
b. Standardization (eg. Gender (F, M), race (AfricanAmerican→ AfAm, Caucasian→ Cauc, Hispanic→Hsp,
c. Removing irrelevant columns (payer code: had 52 % null vals)
d. Colum named→(Glucose serum test) replaced none (not taken) with "0" and normal with "1" , and >200 is replaced with 2 and >300 is replaced with 3
e. Same as point "e" performed the normalization for "**A1c**" test (glycated Hb test).
f. Replaced "NO" of readmitted column with "0".
g. Replaced the "No" with "0" and "Yes" with "1" in diabeticMed column

---

Step4: pivot analysis

Here created various pivot table to get insight form them can be seen in excel sheet.

Step5: Visualization

Created 3 dashboards for covering various factor affecting the re-admission

Step6: Key Insights

1.In Gender wise disease Distribution and males shows higher number of diabetes symptoms.

2. According to "race wise distribution" Caucasian people are more susceptible to diabetes

3. prominence where we can see age group 70-80 shows more number of diabetes cases.

4. **Females** shows more Re-admission rates than **Males** <30.

5. **Readmission** based on A1c test results (race wised) and we can see **Caucasian and African American** people show higher admission rates

6. In case of Admission severity index 1 (*Emergency: Immediate and critical care is required; life-threatening.*) show highest Number of Readmission under 30 days.

7. People who stayed less than a week in Hospital shows the greatest number of re-admissions which somewhere shows inappropriate way of approaching disease treatment.

8. People with severe diabetes means showed the "**Max glucose serum test results"** as 2 & 3 (>200, >300), are mostly readmitted under <30 days.

9. People went through a smaller number of labs procedure shows higher number of readmissions

10. In case of Metformin (***oral antidiabetic medication that helps control blood sugar levels in people with type 2 diabetes.)*** around 70 % people showed the re-admission if it wasn't given , where around 25 % people were also readmitted even if the medication was given at steady level.

11. We do see the greater number of re-admission where the **Repaglinide** (*oral antidiabetic drug that stimulates insulin release to lower blood sugar in type 2 diabetes.)* wasn't given to patients

12. almost similar result can be seen for **Glimepiride** (sulfonylurea that lowers blood sugar by stimulating insulin secretion in type 2 diabetes.)

13. In case of **Glyburide** (*also known as glipalamide-is a sulfonylurea antidiabetic drug that lowers blood sugar by increasing insulin release from the pancreas in type 2 diabetes.)* the result are very similar.

14. We do see a biased result in case of **Insulin** where those who were given the Insulin at steady levels were also re-admitted in huge numbers.

----------End of excel-analysis-report----------