

PREDICTING AND REDUCING HOSPITAL READMISSION USING PATIENT LEVEL DATA

Courtesy:

Mahadi Islam

Department of Statistics, University of Dhaka

Email: phenomenalthe65@gmail.com

Predicting and Reducing Hospital Readmission Using Patient Level Data

1. Abstract and Introduction

Hospital readmissions are a major concern for healthcare system 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 readmission not only improves patient outcomes but also significantly lowers healthcare costs. As our hospital is determined to give away its self for the sake of our patients, high readmission rate is a stigma to our reputation. So, it is high time we took necessary steps to solve this problem as soon as possible and we as a part of it are ever ready to help in ways it expects from us.

2. Hypothesis and Research Questions

Then again, we need to understand why readmissions should be reduced. Our main concern or purpose of this study is to find out;

This study aims to answer the following key questions:

1. Which patient-level factors are most predictive of 30-day hospital readmission among patients with chronic diseases, including diabetes, circulatory, and respiratory disorders?
2. How can insights from patient data be used to design effective, targeted interventions to reduce preventable hospital readmissions and improve overall care quality?

Before moving on with our detailed analyses, our hypothesis would be “Patients admitted with circulatory system diseases, diabetes, or respiratory disorders and those with multiple comorbidities or higher healthcare utilization (such as frequent prior hospitalizations and emergency visits) have a significantly higher likelihood of being readmitted to the hospital within 30 days compared to patients without these risk factors, regardless of age or gender.”

3.Data Description

We are provided with a data set containing some about 28 variables containing a sum total of 1,00,000 observation units but we have decided to work with 14 variables, among which there is our dependent variable ‘readmission status’ and the rest 13 our independent variable which practically affect the readmission status, the variables are;

1. med_change_status
2. procedure_count
3. medication_columns
4. provider_specialty
5. body_weight
6. sex_identity
7. glucose_test_result
8. diagnosis_primary
9. adm_type_code
10. A1C_result
11. ethnic_group
12. diagnosis_secondary
13. insurance_code

we chose these variables because we checked the correlation of these variables with the dependent variable and these came out more likely to have a somewhat meaningful relationship with the dependent variable “readmission_status”.

4.Methods

- This study will utilize a retrospective analysis of 100,000 hospital admissions, focusing on identifying patient-level predictors of readmission within 30 days. We will characterize each admission by demographic variables (such as age and gender), disease group (e.g., circulatory, diabetes, respiratory), comorbidities, count of prior inpatient and outpatient visits, emergency department use, diagnosis count, and medication burden.

- At first, we will perform correlation analyses to identify variables most associated with readmission status. Categorical and continuous variables were compared using cross-tabulation to highlight differences in healthcare utilization and medical complexity between readmitted and non-readmitted patients.
- Then we will construct a binary logistic regression model to assess the combined impact of these predictors on the probability of readmission. Then we will evaluate Model performance using a confusion matrix to determine accuracy, sensitivity, and specificity, and visualize our findings with bar charts and pie charts.
- And finally, we will identify risk factors and patient groups through this approach and use it to formulate targeted recommendations for reducing readmission rates in hospital practice.

This will be our methods and analysis of the whole study.

5. Exploratory Data Analyses

5.1 Descriptive overview

This chart visually compares the mean, median, and standard deviation for key numeric variables in your dataset (e.g., lab test count, procedure count, medication count, diagnosis total).

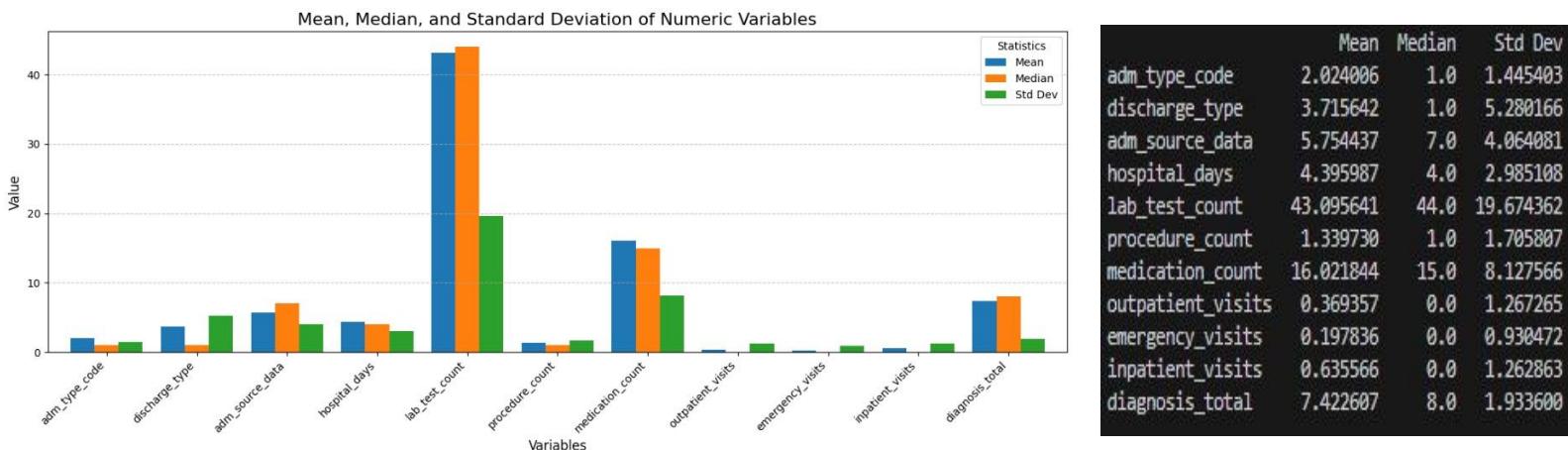


Fig 1: Comparison of The Mean, Median and Standard Deviation

Descriptive analysis showed that most numeric variables, such as procedure count and diagnosis total, had means and medians closely aligned, indicating relatively symmetrical distributions. However, variables like lab test count and medication count had large standard deviations and skewed means versus medians, pointing to notable variability and a small subset of patients with much higher resource use. These findings highlight the diversity and complexity within the analyzed cohort, emphasizing the importance of examining risk by subgroup rather than relying solely on aggregate trends.

5.2 Readmission Rate Overview

Despite 53.9% of total patients have not readmitted thanks to our robust health services but a staggering 46.1% is still getting admitted. This means our health service is still not the ideal one so far.

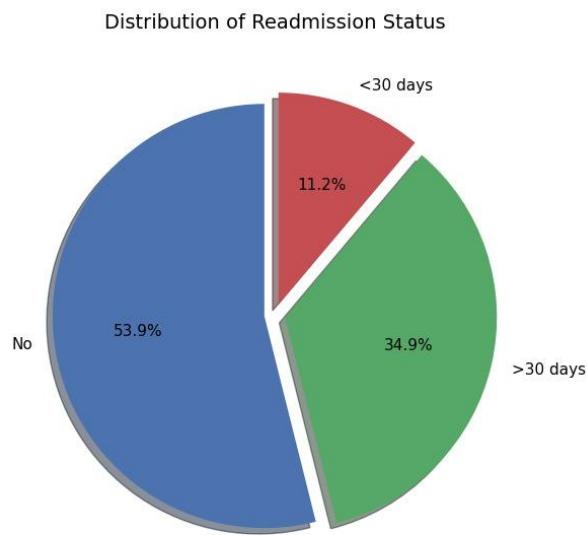


Fig 2: Pie Chart of Readmission Status

Readmission is a bad sign for a hospital. High admission rate may mean-

- poorly coordinated care, inadequate discharge planning, and missed opportunities for preventive intervention.

- inefficient use of hospital resources, requiring extra beds and staff for repeat cases rather than new patients.
- weak or incomplete discharge processes, with patients not fully prepared or supported to manage their health after leaving the hospital.
- patients who are readmitted frequently experience more stress, disruptions, and are at higher risk of complications and further decline in health. Due to frequent readmission, our sole purpose of admission gets ruined because it eventually worsens the condition of the patient both mentally and physically.
- readmission rates serve as a key metric of hospital care, quality and care transitions. Efforts to lower these rates often enhance overall care coordination and patient follow-up.

In a nutshell this can damage the reputation of the hospital very much. So, we are urging that steps are to be taken from this instant.

Below is a line chart demonstrating how 30-day readmission rates for heart attack, heart failure, and pneumonia significantly declined after implementation of the Hospital Readmissions Reduction Program (HRRP) between 2010 and 2016.

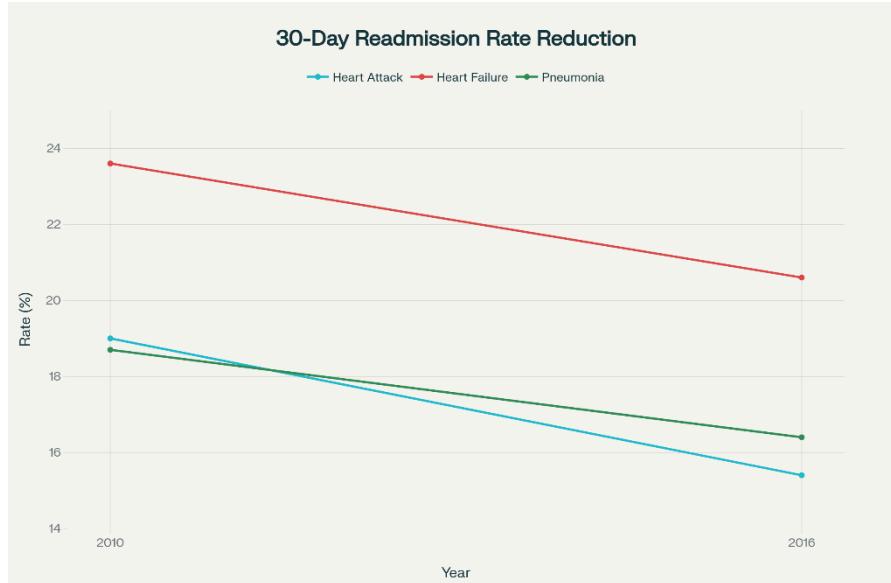


Fig 3: Reductions in 30-Day Readmission Rates after HRRP Implementation (2010 vs 2016)

Source: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6232438/>

After implementation of the HRRP (2010–2016), 30-day readmission rates for targeted conditions dropped significantly:

- Heart attack: from 19.0% to 15.4%
- Heart failure: from 23.6% to 20.6%
- Pneumonia: from 18.7% to 16.4%.

These reductions corresponded to an estimated 150,000 fewer hospital readmissions in just two years (2012–2013), according to analysis by the U.S. Department of Health and Human Services.

To conclude, reducing readmission can change our overall health sector to a great extent.

5.3 Bivariate Analysis and Correlations and Variable Association

Below is an overview of the correlation of all the variables and the chosen ones;

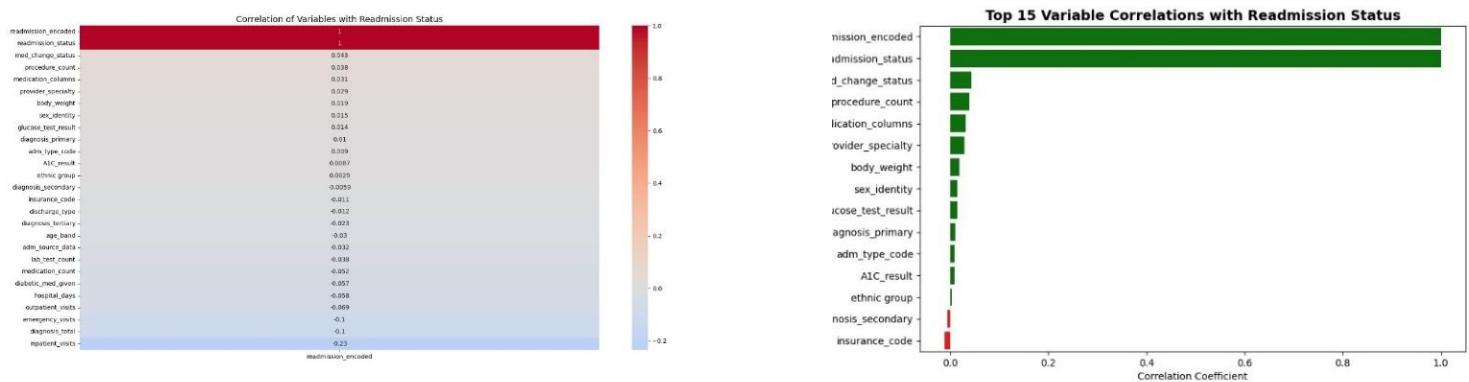


Fig 4: Comparative Correlation Coefficient Diagram

There are some notable insights that we learned from the correlation of these variables, such as;

- The strongest negative correlation was found with **inpatient visits** (**-0.23**), meaning more inpatient visits are associated with a lower likelihood of readmission, possibly reflecting closer monitoring or effective treatment during inpatient stays.
- Other variables with weak negative correlations include **diagnosis total** (**-0.1**), **emergency visits** (**-0.1**), and **outpatient visits** (**-0.069**), suggesting that higher use of these services is modestly linked to lower readmission rates.

- Positive but weak correlations were observed with factors like **medication change status** (0.043), **procedure count** (0.038), and **number of medication columns** (0.031), indicating that patients experiencing medication adjustments or more procedures may have a slightly higher risk for readmission.
- Most categorical and demographic variables (such as sex, age, and insurance code) show almost no correlation, implying little direct impact on readmission status individually.
- To dive further into our analysis, we ran a crosstab and regression analysis. Let's uncover what they have to say;

Different factors affecting Readmission status:											0	1	
readmission_status	inpatient_visits	med_change_status	diabetic_med_given	diagnosis_total	emergency_visits	outpatient_visits	hospital_days	medication_count	lab_test_count	procedure_count			
0	0	0	1	0	0	1	1	1	0	0	1.0	0.0	
									41	0	1.0	0.0	
									3	32	0	1.0	0.0
									4	47	0	1.0	0.0
									5	39	0	1.0	0.0
...											
17	0	0	1	0	0	1	9	34	0	0	0.0	1.0	
18	0	0	9	9	1	7	24	104	5	0	0.0	1.0	
19	1	1	6	0	0	8	8	37	1	0	0.0	1.0	
			9	13	0	3	16	63	0	0	0.0	1.0	
21	0	1	7	1	0	1	10	38	0	0	0.0	1.0	

Fig 5: Crosstab Analysis

The crosstab table displays the relationship between readmission status and several independent variables, such as inpatient visits, medication changes, diagnosis count, emergency/outpatient visits, and counts of days and procedures.

Differences in average values of these predictors between readmitted and non-readmitted groups can help identify risk factors. For example:

- Higher total diagnoses and more frequent emergency/outpatient visits may be associated with higher readmission.
- Medication count and lab/test count also appear higher in some readmitted groups, suggesting more complex care needs.

To sum up, These patterns highlight that patients with more medical complexity, care touchpoints, or acute events are at greater risk for readmission.

5.4 Group Comparison

Below we can see a boxplot that visualizes and compares the distributions of multiple variables such as medication count, inpatient visits, outpatient visits, emergency visits, diagnosis total, and A1C result across readmission status groups (No readmission, readmitted after 30 days, and readmitted within 30 days).

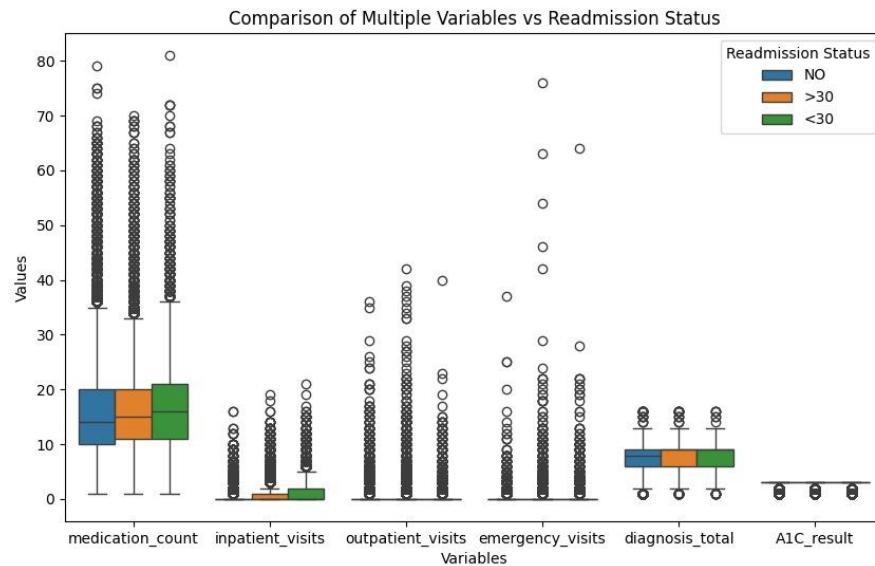


Fig 6: Boxplot to Compare Various Variables.

- **Medication Count:** Patients readmitted (both <30 and >30 days) generally have higher median and upper-range medication counts than those not readmitted, suggesting a relationship between medication burden and readmission risk.
- **Inpatient Visits:** Higher frequency and upper-range (outlier) counts among readmitted patients, especially within 30 days.

- **Outpatient/Emergency Visits:** Both types are elevated in readmitted groups compared to non-readmitted, with substantial variability (many high outliers), reflecting higher healthcare utilization in those at greatest risk.
- **Diagnosis Total:** Readmitted patients have greater total diagnoses on average, again indicating more medical complexity.
- **A1C Result:** Appears fairly evenly distributed with minimal difference across readmission groups, suggesting glycemic control has less direct effect here.

This boxplot reinforces that hospital readmission is strongly associated with higher counts of medications, inpatient and outpatient encounters, emergency visits, and overall diagnosis burden. These patterns are particularly evident in the group readmitted within 30 days. The spread (range) of these variables is broad, with many outliers, showing that while most readmitted patients have moderately high values, a small subset have extreme utilization.

5.5 Regression Model Performance and Predictive Patterns

The regression model (confusion matrix) which is binary classification regression is used to predict a patient will be readmitted or not, based on multiple patient-level factors.

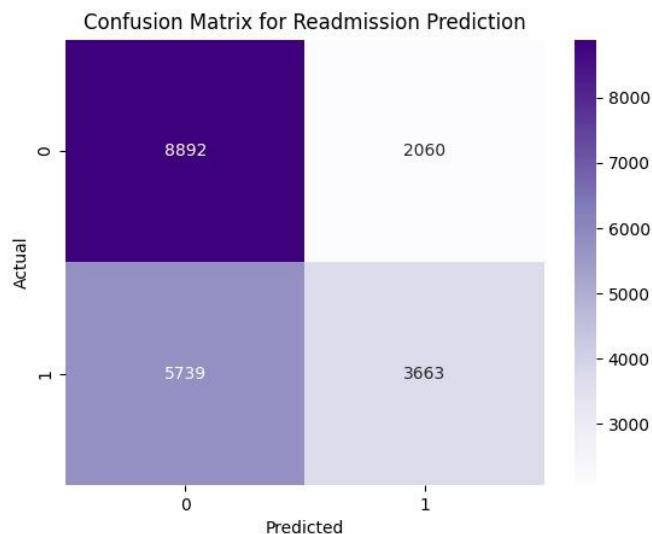


Fig 7: Model for Readmission Prediction

Now let's see what we can learn from our model:

- The model successfully identifies many patients who are not at risk for readmission (high true negatives), suggesting our data includes strong signals for low-risk cases.
- However, the large number of false negatives (actual readmissions missed by the model) highlights limitations in detecting all high-risk individuals.
 - 8,892 true negatives: not readmitted and model correctly predicted not readmitted.
 - 3,663 true positives: readmitted and model correctly predicted readmitted.
 - 2,060 false positives: not readmitted but model incorrectly predicted readmission.
 - 5,739 false negatives: readmitted but model incorrectly predicted no readmission.
- This pattern across our model and data demonstrates that while certain risk features (diagnosis count, frequent emergency visits, higher medication/lab counts) are associated with readmission, no single variable is a reliable predictor alone. Our correlation analysis supports this, as most variables only weakly relate to readmission.

Despite its success the model is not flawless, it lacks in significant areas like:

- The model misses a significant number of true readmissions (high false negative), indicating poor detection of high-risk patients.
- It is much better at predicting who will not be readmitted than who will be, which is a limitation since preventable readmissions are the key concern.

Overall model assesment:

Model Performance Metrics	
Accuracy	0.617
sensitivity	0.390
Specificity	0.812
F1 Score	0.484

6. Key Findings and Figures

As we are done with our model we would like point out our key findings regarding our analysis and we shall see what steps are needed to be taken to improve our overall problems.

1. None of the variables (other than the encoded status itself, used for reference) approached even moderate correlation values. This suggests that readmission risk is multifactorial and cannot be reliably predicted by any single patient characteristic or service utilization metric alone.
2. Readmission is a complex, multifactorial outcome in your dataset: the absence of strong predictors means interventions must draw from combined risk factors, not isolated variables.
3. The model reflects real-world challenges in hospital readmission science—many factors (clinical, social, behavioral) interact, and not all are captured in structured data.
4. Even with a large dataset (100,000+ cases), decision-making for high-risk patients requires both quantitative modeling and clinical judgment.

We would also like to point a quite interesting take on our readmission status.

Below we are witnessing a heat map of readmission status by specific disease groups.

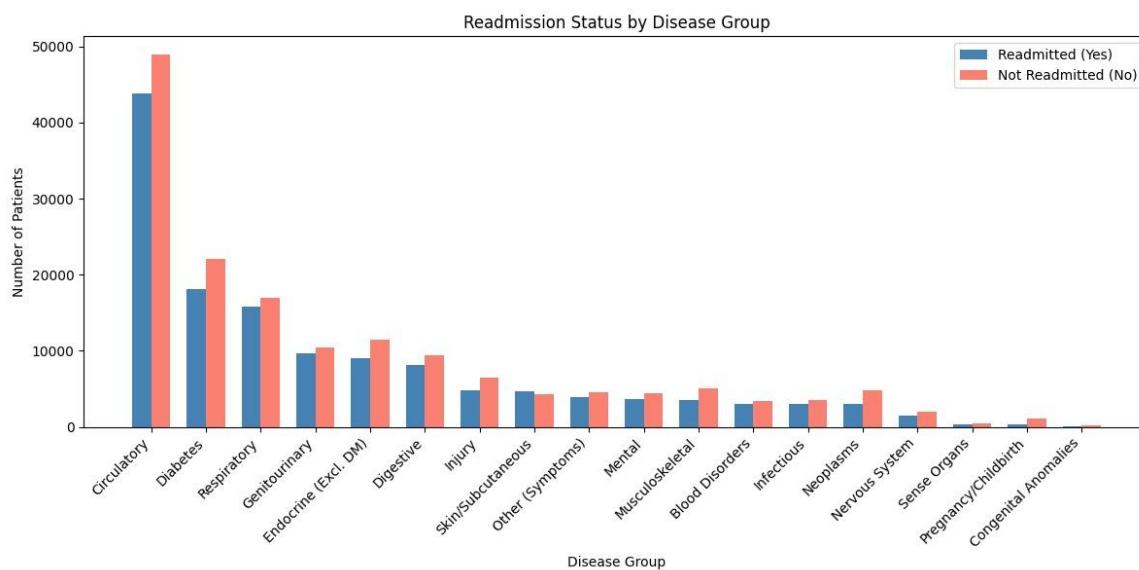


Fig 8: Readmission Status by Disease Group

This chart compares the number of patients with and without hospital readmission, categorized by disease group. The disease groups with the highest patient counts (readmitted and not readmitted) are **Circulatory**, **Diabetes**, and **Respiratory** disorders. Less frequent disease categories like Sense Organs, Pregnancy/Childbirth, and Congenital Anomalies have much lower overall readmission numbers. The data shows that patients with circulatory and diabetes-related diagnoses are much more represented in both readmission and non-readmission groups, indicating these are major contributors to hospital utilization and should be a focus for targeted interventions.

If this one piqued our interest, we should look at another graph.

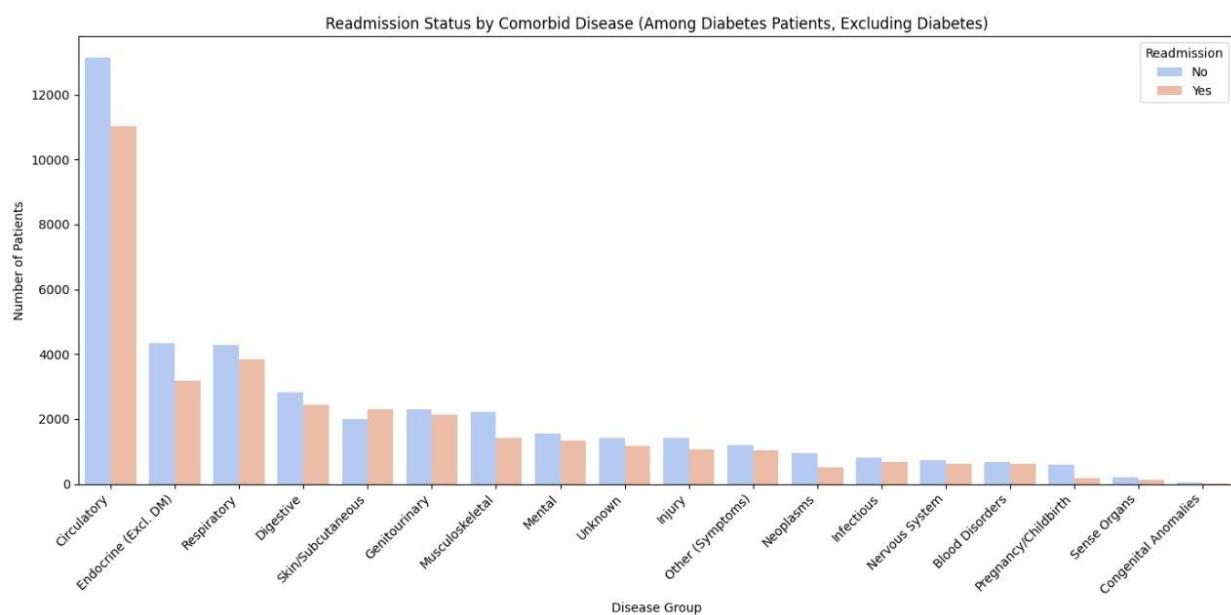


Fig 9: Readmission Status by Comorbid Disease Among Diabetes Patient.

This chart focuses specifically on diabetic patients and examines readmission rates based on their comorbid disease group, excluding diabetes itself. As with the previous chart, **Circulatory** comorbidities are the most common in both readmitted and not readmitted diabetic patients. Next most prevalent are **Endocrine (excluding DM)**, **Respiratory**, and **Digestive** comorbidities. The overall pattern is similar to the full population, but with even more dominance of circulatory and endocrine comorbidities among diabetic populations. This chart suggests that among diabetic patients, those with additional circulatory or other endocrine

disorders are particularly at risk for hospital readmission and should be prioritized for enhanced monitoring and transitional care.

7. Discussion and Conclusion

Our intentions of presenting this data is that, while readmitting a certain patient what metrics are we focusing on? This can be severity of patients condition or does that patient have any other disease, diagnosis and risk level, patient-level and social factors etc. If we look into these two graphs alone, certain disease have high readmission rate and that raises a question of why?

Aren't we treating those disease with the same level of care that we are putting behind diseases with lower readmission rate or are they false readmission? We need to look into this matter ASAP.

Now to focus on the big question of what measures do we take?

1. Firstly we target high risk disease groups.
 - a. We should focus interventions on patients with **circulatory, diabetes, and respiratory** diseases, as these categories have the highest readmission counts.
 - b. Same attention is needed for diabetic patients with additional **circulatory or endocrinological disorders**.
2. Secondly, we enhance discharge planning and transitional care.
 - a. We must implement structured discharge processes, including medication review, clear instructions, and scheduling prompt follow-up appointments.
 - b. We can assign care transition teams or coordinators for patients with complex conditions to help manage their outpatient needs and follow-up after discharge.
3. Thirdly, we must emphasise medication management and education.
 - a. We must ensure that patients have a thorough understanding of their medication because many readmission stem from drug interactions or confusion about new regimens.
 - b. We can offer individualised plans for diabetes and multi-comorbidity patients.

4. Finally, we must address social determinants of health.
- We can assess for functional impairment, inadequate home support, low health literacy, and transportation barriers all of which increase readmission risk.
 - We can provide support services (social work, community care coordination) for those most vulnerable.

To give a birds eye view;

Solution	Who to target	Data insight source
Enhanced transitions	High diagnosis/procedure/med count	Regression/Crosstab
Focused care for top diseases	Circulatory, diabetes, respiratory	Disease group bar charts
Medication reconciliation	Multiple comorbidities	Model and readmission data
Early follow-up	High-risk patients post-discharge	Correlation, regression
Address social barriers	Patients with poor support	Admission assessment factors

Applying these interventions systematically, guided by our data analyses and model insights, we can measurably reduce avoidable readmissions and improve both patient and hospital outcomes.

8. Tools

Python programming language and additional libraries (Numpy, Matplotlib, Pandas, Seaborn, Scipy-learn).