Framing the Analysis

## **Thoughts**

1. Find your problem statement, not use your data to create the problem statement.

## **Quotes**

1. “Without data, you’re just another person with an opinion.” William E. Demings
2. “If you torture the data long enough, (nature) it will confess.” Ronald H. Coase

## **References**

[35 Healthcare KPIs to Track in 2025 | NetSuite](https://www.netsuite.com/portal/resource/articles/erp/healthcare-kpis.shtml)

[18 Healthcare KPIs to Measure Hospital Performance | Medium](https://houseofanalytics.medium.com/18-healthcare-kpis-to-measure-hospital-performance-4896b8cb0d5a)

[Big Healthcare Data Analytics: Challenges and Applications](https://www.comp.nus.edu.sg/~ooibc/healthcarechapter.pdf)

[2025 global health care outlook | Deloitte Insights](https://www.deloitte.com/us/en/insights/industry/health-care/life-sciences-and-health-care-industry-outlooks/2025-global-health-care-executive-outlook.html)

[Lewis, M. N., & Nguyen, T. H. (2020). *Transforming Healthcare Analytics: the quest for healthy intelligence*. Wiley.](https://www.amazon.sg/Transforming-Healthcare-Analytics-Healthy-Intelligence/dp/111961354X)

[What is a Patient Services Representative and Why is this Role Critical in Healthcare?](https://info.pmimd.com/blog/what-is-a-patient-services-representative-and-why-is-this-role-critical-in-healthcare)

[Hospital Operations Management: Definition and Career Paths |](https://www.indeed.com/career-advice/career-development/operations-hospital) [Indeed.com](http://indeed.com)

[Effective bed management system at Singapore General Hospital | SMU Newsroom](https://news.smu.edu.sg/news/2024/11/01/effective-bed-management-system-singapore-general-hospital-0)

[Predictive Analytics Tool Identifies Readmission Risk, Reduces Costs | TechTarget](https://www.techtarget.com/healthtechanalytics/news/366590437/Predictive-Analytics-Tool-Identifies-Readmission-Risk-Reduces-Costs)

[Predictive Analytics Use EHR Data for Hospital Readmissions | TechTarget](https://www.techtarget.com/healthtechanalytics/news/366590523/Predictive-Analytics-Use-EHR-Data-for-Hospital-Readmissions)

[Systematic Review and Meta-Analysis of the Financial Impact of 30-Day Readmissions for Selected Medical Conditions: A Focus on Hospital Quality Performance](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011876/)

[Potentially avoidable readmissions: Understanding drivers and technology-enabled solutions - Annals Singapore](https://annals.edu.sg/potentially-avoidable-readmissions-understanding-drivers-and-technology-enabled-solutions/#:~:text=Hospital%20admissions%20places%20high%20resource%20demands%20on%20the,%2830-day%20readmissions%29%20remains%20imperative%20to%20ensure%20long-term%20sustainability)

[Evaluating Classification Models: A Guide to Choosing the Right Metrics | by Chaw Thiri San | Medium](https://medium.com/@chawthirisan/evaluating-classification-models-a-guide-to-choosing-the-right-metrics-66f7a75a3b52)

[Frequent hospital admissions in Singapore: clinical risk factors and impact of socioeconomic status - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC5778257/)

<https://www.nus.edu.sg/newshub/news/2017/2017-03/2017-03-28/HOSPITAL-st-28mar-pB2.pdf>

Acknowledgements:

Beata Strack, Jonathan P. DeShazo, Chris Gennings, Juan L. Olmo, Sebastian Ventura, Krzysztof J. Cios, and John N. Clore, “Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records,” BioMed Research International, vol. 2014, Article ID 781670, 11 pages, 2014.

## 

## **Introduction**

Hospital readmission rate is one of the key performance metrics in healthcare, showing both clinical quality and operational efficiency. It is normally measured as the percentage of patients who return to a hospital, staying overnight, within a specific timeframe, normally 30-days after discharge.

According to a study published in the National Library of Medicine in Mar 2024, the estimated average cost of a 30-day all-cause adult readmission is at USD 16,037.08.1 Reducing the readmission rate or intervening before a patient is readmitted is beneficial both to hospitals and, more importantly, patients by improving outcomes and reducing avoidable complications.

1. [Systematic Review and Meta-Analysis of the Financial Impact of 30-Day Readmissions for Selected Medical Conditions: A Focus on Hospital Quality Performance](https://pmc.ncbi.nlm.nih.gov/articles/PMC11011876/)

## **Problem Statement**

This project aims to predict whether a patient will be readmitted after discharge from the hospital. (Assumption: readmitted label is defined as 30-day after discharge from the hospital based on the original data source)

By modeling readmission risk using features such as age, time in hospital, diagnoses, hospital visits, and procedures done on the patient, this project seeks to support hospitals in identifying high-risk patients and improving care pathways.

## **Stakeholders involved**

[**Case Managers**](https://www.casemanagement.sg/about-us/about-case-manager/what-is-a-case-manager/)

[**Operations Managers**](https://www.indeed.com/career-advice/career-development/operations-hospital)

**Clinicians**

[Ambulatory](https://everynurse.org/what-is-ambulatory-care-nursing/) / Outpatient / Inpatient / Emergency Nurse

[**Patient Services Representatives**](https://info.pmimd.com/blog/what-is-a-patient-services-representative-and-why-is-this-role-critical-in-healthcare)

**Clinicians, nurses, case managers, and hospital administrators** can use readmission risk predictions to guide targeted interventions — such as timely follow-up care (e.g. Hospital-to-Home programme), and proactive care coordination — all aimed at improving patient outcomes and optimizing healthcare spending.

1. [Hospital to Home | SingHealth](https://www.singhealth.com.sg/about-singhealth/who-we-are/community-population-health/population-health/hospital-to-home#:~:text=Hospital%20to%20Home%20%28H2H%29%20is%20a%20nationwide%20programme,social%20needs%2C%20and%20a%20high%20risk%20of%20readmissions).

## 

## **Objectives / Goals**

1. Analyze the readmission pattern across the age group, diagnosis types, and procedures.
2. Investigate if diabetes as primary diagnosis correlates strongly with readmission.
3. Build & select a predictive model to estimate the readmission of a patient.

## **Success Metrics**

1. Identify the patterns in readmission across the age group, diagnosis types, and procedures.
2. Conduct statistical tests to assess correlation between diabetes diagnosis and readmission
3. The predictive model achieves a strong F1 score to balance precision and recall.

## **Analysis Approaches**

1. Develop hypotheses for the analysis.
2. Data Collection.
3. Perform Exploratory Data Analysis to know the data. As part of this analysis, perform the:
   1. Descriptive Statistic Analysis,
      1. Identify missing values, evaluate the way to handle the missing values like imputation
      2. Check for outliers
   2. Diagnostic Analysis,
   3. Predictive Model Analysis,
      1. Use *information gain* to select an important feature.
      2. When there is missing data, evaluate its pattern with the target response to decide the approach to handle it.
      3. Evaluate the bias of the dataset
      4. Build a logistic regression model
      5. Evaluate the model:
         1. Confusion matrix
   4. Features definition
4. Perform where required:
   1. data cleaning
   2. feature engineering
5. Define, build, train, and test the predictive model.
6. Storyboard the analysis.

## **Risks**

1. Changes in the healthcare landscape like treatment protocols and hospital policies over time could affect the readmission rate.
2. Implementing the predictive model into the hospitals’ current workflow and ensuring its adoption can be challenging.

## **Assumptions**

1. The dataset of 25,000 records is representative of the broader population of the hospital.
2. The features in the dataset are relevant and sufficient for predicting patient readmission.
3. The dataset is reasonably correct, without widespread systematic errors or critical missing values that would skew our analysis.
4. The ‘Missing’ category in the medical\_specialty feature is a valid and useful signal for our Logistic Regression predictive model.
5. The factors that influence patient readmission remain relatively stable over the time period covered in the dataset.

## **Datasets**

1. <https://app.datacamp.com/learn/competitions/hospital-patient-readmissions>
   * [Diabetes 130-US Hospitals for Years 1999-2008 - UCI Machine Learning Repository](https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008)