PROJECT REPORT / DOCUMENT OUTLINE

TOPIC: Diabetes Patient Readmission Analysis

Sector: Healthcare Analytics

Team Members Details:

Name	Enrolment Number	Role
Mitul Bhatia	2401010279	Project Lead
Ramani Dhruv Dineshbhai	2401010153	Analysis Lead
Divya Singh	2401020020	Data Lead
Anushka Tyagi	2401010090	PPT & Quality Lead
Daniel Tayal	2401010140	Strategy Lead
Aaryan Gera	2401010009	Dashboard Lead

1. Executive Summary

Hospital readmissions within 30 days are a major performance and cost indicator in healthcare systems. Frequent readmissions increase hospital burden, raise costs, and indicate possible gaps in discharge planning or disease management.

This project analyses the **Diabetes 130-US Hospitals dataset (1999–2008)** to identify key drivers of 30-day readmission. The goal was to understand which clinical and demographic factors are associated with higher readmission risk and to translate those findings into actionable hospital-level recommendations.

Using Google Sheets for cleaning, transformation, pivot analysis, and dashboard development, we:

- Reduced raw data (~101,766 rows) to **14,116 unique patients**
- Built binary risk flags
- Created ICD-9 condition groupings
- Estimated financial impact of readmissions
- Designed an interactive dashboard

Key Insights

- Overall 30-day readmission rate: **13.2%**
- Emergency admissions show highest readmission rate (14.1%)
- 83.1% of patients had no A1C test recorded
- 94.9% had no max glucose test
- Emergency admissions account for **$12.3M of $20.49M total estimated cost**

Key Recommendations

- Standardize A1C testing at discharge
- Strengthen emergency discharge protocols
- Flag long hospital stays (>4.5 days) before discharge
- Prioritize elderly and high-risk patients for follow-up

This analysis provides a structured risk-identification framework that can support hospital administrators in reducing avoidable readmissions.

2. Sector & Business Context

Sector Overview

The healthcare sector continuously monitors 30-day readmission rates as a quality and financial performance indicator. In the United States, high readmission rates may lead to financial penalties under reimbursement models.

Current Challenges

- Poor chronic disease monitoring
- Incomplete discharge planning
- Inconsistent follow-up systems
- High emergency admission burden

Why This Problem Was Chosen

Diabetes is a chronic condition requiring continuous management. The dataset provides a strong opportunity to evaluate readmission drivers in a structured manner.

---

3. Problem Statement & Objectives

Problem Statement

To identify key demographic, clinical, and administrative factors that influence 30-day hospital readmission among diabetic patients.

Project Scope

- Analyze first hospital encounter per patient
- Focus on 30-day readmission
- Perform data cleaning and transformation in Google Sheets
- Build interactive dashboard
- Estimate financial impact

Success Criteria

- Accurate readmission rate calculation
- Clear KPI framework
- Identifiable high-risk segments
- Business-level recommendations supported by data

---

4. Data Description

Dataset Source

Diabetes 130-US Hospitals Dataset
UCI Machine Learning Repository
Original Period: 1999–2008
https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008

Raw Dataset Size

- ~101,766 rows
- 50+ columns

Final Analytical Dataset

- 14,116 rows
- Single encounter per patient
- Cleaned and structured and randomized

Data Structure

The dataset includes:

- Patient identifiers
- Demographics
- Admission details
- Clinical test results
- Medication information
- Readmission outcome

Key Limitation

- Data is historical (1999–2008)
- No BMI data (weight dropped due to 97% missing)
- No socioeconomic variables

---

5. Data Cleaning & Preparation

All cleaning was performed in **Google Sheets**, as required.

Single Encounter Rule

Multiple visits existed per patient.
We sorted by patient_nbr and encounter_id and retained only the first encounter.

Missing Value Handling

- race: 2,273 "?" replaced with mode (Caucasian)
- gender: 3 invalid rows deleted
- medical_specialty: "?" replaced with "Missing"
- weight: dropped (>97% missing)
- payer_code: dropped (>40% missing)

Column Reduction

- 24 medication columns removed
- And other no involved columns

Feature Engineering

Created derived columns:

- median_age
- Primary_Condition (ICD-9 grouping)
- A1C_None
- A1C_High
- max_glu_high
- Readmission Rate (binary)
- High_Risk_Flag

- Admission_Group
- Medication_Status
- Total Cost

All formulas documented in Data Dictionary.

---

6. KPI & Metric Framework

KPI 1 — 30-Day Readmission Rate

Formula:
Readmission Rate = (Readmitted <30 days) / Total Patients

Result: 13.2%

KPI 2 — Untested A1C Rate

Formula:
A1C_None = 1 if test not performed

Result: 83.13%

KPI 3 — High Risk Flag

Patients with:

- Readmission = 1
  AND
- A1C_High = 1

Result: 7.2%

KPI 4 — Estimated Financial Impact

Assumption:
Average cost per readmission = $11,000

Total cost = 1,863 × 11,000
= **$20,493,000**

---

7. Exploratory Data Analysis (EDA)

Admission Type Analysis

Emergency admissions:

- 56% of total
- 14.1% readmission rate
- Highest cost contribution

Age vs Readmission

Highest readmission seen in age group 80–90 (14.7%).

Hospital stay increases with age.

Length of Stay

- Non-readmitted: 4.27 days
- Readmitted: 4.78 days

Longer stay signals higher risk.

A1C Testing Gap

83.1% of patients were not tested.

Indicates process gap in chronic disease monitoring.

Medication Change Impact

- Stable medication: 12.2% readmission
- Changed medication: 14.3% readmission

Medication instability correlates with readmission.

---

8. Advanced Analysis

ICD-9 Grouping

522 primary diagnosis codes grouped into:

- Diabetes
- Circulatory
- Respiratory
- Other

Circulatory conditions show the highest readmission share.

High Risk Segmentation

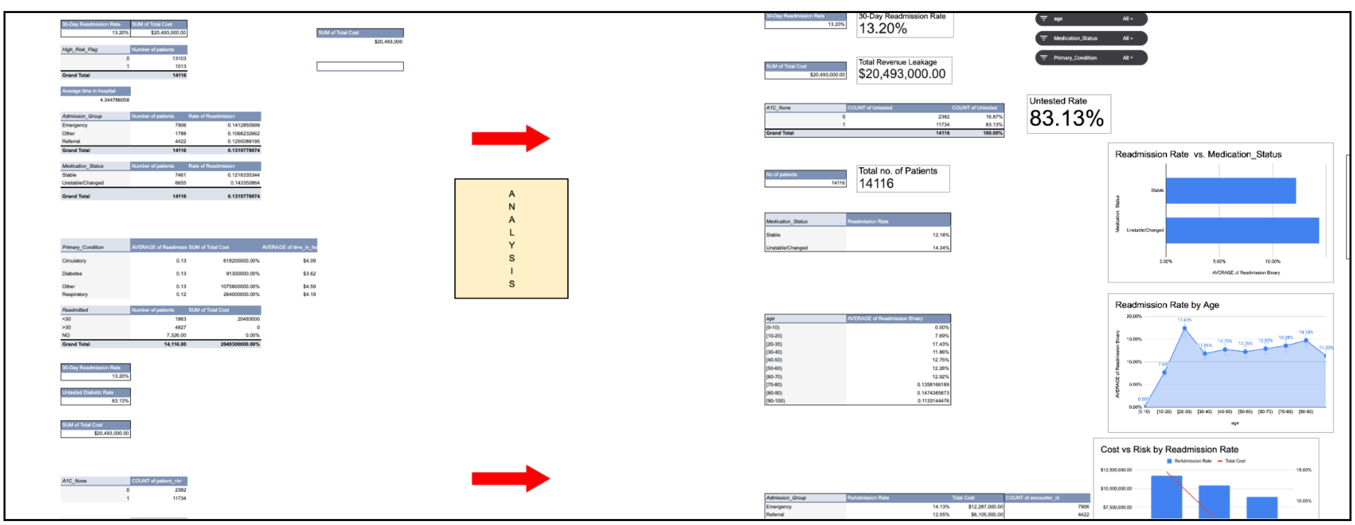
High_Risk_Flag identified 1,013 patients
(7.2% of dataset)

These patients represent the highest ROI intervention group.

Cost Breakdown by Admission Group

Emergency admissions account for:

$12.28M of total $20.49M



9. Dashboard Design

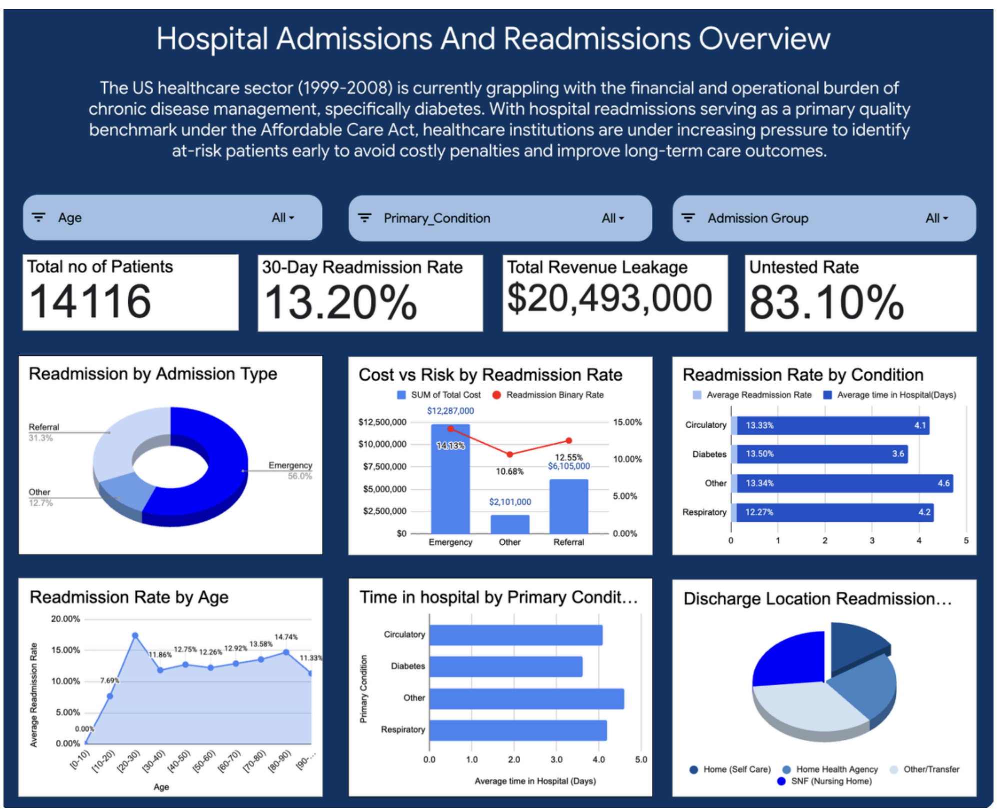
Developed in Google Sheets using:

- Pivot tables
- Binary flags
- Conditional formatting
- Interactive filters

Dashboard Objective

Provide hospital administrators with:

- Readmission overview
- Cost impact
- Risk segmentation
- Filterable demographic insights



Filters Available

- Age group (0-100)
- Primary condition (Circulatory , Diabetes , Respiratory , other)
- Medication status (Emergency , Referral , other)

---

10. Insights Summary

1. 13.2% overall readmission rate
2. Emergency pathway drives highest cost
3. 83% A1C testing gap
4. 94.9% glucose test gap
5. Older age increases risk
6. Longer stays increase risk
7. Medication instability increases risk
8. Circulatory conditions contribute significantly
9. 7.2% patients are high-priority intervention group
10. $20.49M estimated cost impact

11. Recommendations

1. Standardize A1C Testing

Addresses 83% testing gap.

2. Emergency Discharge Checklist

Focus on the highest cost pathway.

3. Auto-Flag Long Stay Patients

Stay > 4.5 days triggers review.

4. High-Risk Follow-Up Program

Target High_Risk_Flag patients first.

Each recommendation has high feasibility and moderate implementation cost.

---

12. Impact Estimation

If readmission reduces by 5%:

Savings ≈ $1–2M annually (scaled hospital estimate)

Benefits:

- Reduced penalties
- Better bed availability
- Improved patient outcomes

---

13. Limitations

- Historical dataset (1999–2008)
- Cost assumption fixed at $11,000
- No causal inference (correlation only)
- No socioeconomic data

---

14. Future Scope

- Predictive ML model
- Logistic regression risk scoring
- External validation on modern data
- Integration with EMR systems

---

15. Conclusion

This project successfully:

- Cleaned and structured complex healthcare data

- Identified readmission risk drivers
- Quantified financial impact
- Developed actionable dashboard insights

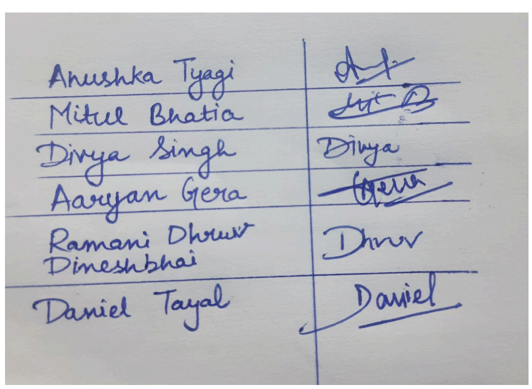
The framework can support hospital administrators in prioritizing high-risk patients and reducing avoidable readmissions.

---

16. Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Mitul Bhatia	✔	✔	✔	✔	✔	✔	Project Lead
Divya Singh	✔	✔	✔		✔		Data Lead
Dhruv Ramani	✔	✔	✔	✔			Analysis Lead
Aaryan Gera	✔		✔	✔			Dashboard Lead
Anushka Tyagi		✔	✔		✔	✔	PPT & Quality Lead
Daniel Tayar		✔	✔	✔			Strategy Lead

Declaration: We confirm that the above contribution details are accurate and verifiable.



Team Signature:
