

# Predicting Hospital Readmissions Using Machine Learning

---

*A Data Science Approach to Reducing Healthcare Costs*

**Presented by:** [Aman Sharma]

# Introduction

---

- **Problem Statement:** Hospital readmissions within 30 days increase costs and risks.
- **Goal:** Build a predictive model to identify high-risk patients.
- **Business Impact:** Reducing readmissions improves patient care and hospital efficiency.



# Dataset Overview

---

- **Data Sources:** Hospital patient records.
- **Key Features:**
  - Demographics: Age, Gender
  - Hospitalization Details: Admission Type, Visits
  - Medical History: Diagnoses, A1C Result, Medications
- **Target Variable:** Readmitted (Yes/No)

# Methodology

---

## 1. Data Preprocessing

1. Handled missing values (median/mode imputation).
2. Encoded categorical variables (One-Hot Encoding).
3. Scaled numerical features.
4. Removed outliers from the 'Age' feature.

---

- **Feature Engineering**

- Created Age Groups.
- Aggregated hospital visit counts.
- Identified high-diagnosis patients.

- **Model Selection & Training**

- Algorithms used: Random Forest, KNN, Logistic Regression.
- Split dataset (70% training, 30% testing).



# Model Performance Metrics

---

- **Random Forest:**

- Accuracy: 50%
- ROC-AUC: 49%

- **KNN:**

- Accuracy: 49%
- ROC-AUC: 50%

- **Logistic Regression:**

- Accuracy: 51%
- ROC-AUC: 52%

# Feature Importance

---

- **Top Predictors of Readmission:**
  - Number of inpatient visits.
  - Total number of diagnoses.
  - Admission type (Emergency admissions had higher readmission rates).

# Insights & Recommendations

---

- **Proactive Patient Monitoring:** Extra post-discharge care for high-risk patients.
- **Targeted Interventions:** Special follow-up programs for chronic conditions.
- **Improved Discharge Planning:** Clearer patient instructions and scheduled check-ups.