

# TASK 2

1. Perform **Exploratory Data Analysis (EDA)** to understand feature distributions and correlations.
2. Preprocess the data (e.g., scaling, encoding if necessary).
3. Train models: Logistic Regression, Random Forest, XGBoost.
4. Evaluate using accuracy, precision, recall, and F1-score.

## *Key Insights:*

1. **Performance Comparison:**
  - XGBoost and Random Forest models outperform Logistic Regression in predicting readmissions.
  - XGBoost achieved the highest **ROC-AUC score**.
2. **Feature Importance:**
  - XGBoost and Random Forest provide insights into feature importance. Key features like age, cholesterol levels, and other patient vitals strongly influence predictions.
3. **Recommendations:**
  - Deploy the XGBoost model for identifying high-risk patients.
  - Focus on optimizing critical patient features to reduce readmissions.
  - Future improvements: Include more features like treatment duration, comorbidities, and demographics for better predictions.

## *Potential Use Cases:*

- Hospitals can use this model to allocate resources efficiently for high-risk patients.
- Improved patient follow-ups based on model predictions to reduce readmission rates.

## Model Results Summary

### *Logistic Regression*

- **Accuracy:** ~81%
- **ROC-AUC:** ~0.93
- **Confusion Matrix:**
  - True Negatives: 70
  - False Positives: 30
  - False Negatives: 9
  - True Positives: 96
- **Insights:** Logistic Regression provides a good balance of precision and recall, especially for identifying positive cases (class 1).

### *Random Forest*

- **Accuracy:** 100%
- **ROC-AUC:** 1.0
- **Confusion Matrix:**
  - Perfect classification on both classes.
- **Insights:** The Random Forest model shows perfect performance, which might indicate overfitting. Further testing on unseen data is advised.