Model & Data Overview

Introduction:
This classification task predicts student pass/fail outcomes based on demographic and academic features from the UCI Student Performance dataset. The goal is to build a robust predictive model using supervised machine learning.

Dataset Summary:

- 395 student records
- 33 features (both categorical and numerical)
- Target: pass (1) if final grade (G3) ≥ 10; otherwise, fail (0)
- Class ratio: Pass ~67%, Fail ~33%

Preprocessing & EDA:

- Performed label binarization on the pass column
- Dropped G1, G2, G3 (raw grades) columns
- Encoded categorical variables with one-hot encoding
- Scaled numerical features using StandardScaler
- Explored feature correlations and class distributions

Model Evaluation & Deployment

Model Performance:

Model	**Accuracy**	**Precision (Pass)**	**Recall (Pass)**	**F1 Score (Pass)**
Logistic Regression	**0.71**	**0.74**	0.87	**0.80**
Random Forest	0.66	0.68	**0.90**	0.78

- **Best model:** Logistic Regression due to its higher overall F1 score and balanced precision/recall

Deployment Strategy Recommendations:

- Export model using Pickle
- Serve predictions via an API
- Deploy on cloud platform

Monitoring Recommendations:

- Implement logging and error tracking
- Set up alerts on model accuracy drops or unusual inputs
- Schedule regular model retraining and validation