Adult Income Prediction Project

 # Results Summary
Model Performance Comparison

Model	Accuracy	ROC AUC	F1 Score (>50K)	Training Time
Random Forest	86.69%	0.9205	0.6953	~2 minutes
XGBoost	87.53%	0.9311	0.7219	~1.5 minutes

 # Best Parameters
Random Forest:
{
'max_depth': 20,
'min_samples_leaf': 2,
'min_samples_split': 2,
'n_estimators': 200
}

XGBoost:
{
'colsample_bytree': 0.8,
'learning_rate': 0.1,
'max_depth': 6,
'n_estimators': 200,
'subsample': 1.0
}

 # Key Findings

Performance:
XGBoost slightly outperformed Random Forest (87.53% vs 86.69% accuracy)
Both models show excellent discrimination (ROC AUC > 0.92)
Better at identifying lower incomes (94% recall) than higher incomes (63-67% recall)

Top Predictive Features:
Age (+0.32 SHAP value range)
Education level (+0.28)
Capital gains (+0.25)
Hours per week (+0.19)
Marital status (+0.17)
Confusion Matrix Insights:

Random Forest:
Correctly identified 94% of <=50K incomes
Correctly identified 63% of >50K incomes

XGBoost:
Correctly identified 94% of <=50K incomes
Correctly identified 67% of >50K incomes