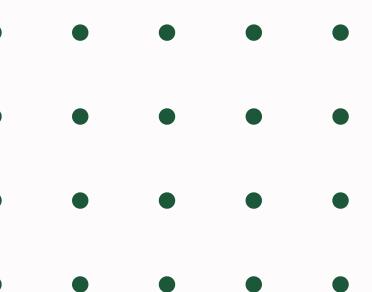




Bank Of India

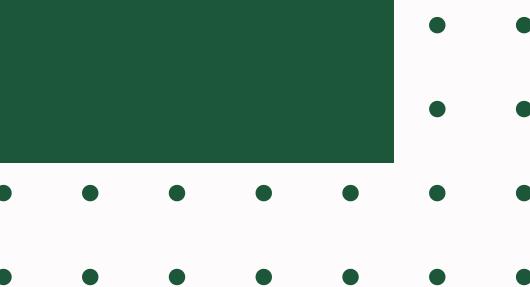
Presentation 2024

LOAN APPROVAL PROJECT



01 Overview
02 Business Understanding
03 Data Understanding
04 Exploratory Data Analysis
05 Models and Evaluation
06 Recommendation
07 Contact

CONTENT



Overview

This project implements machine learning algorithms to analyzes data from previous clients to identify patterns that can be used to flag an applicant as a risky applicant or

OBJECTIVES

01



To create a machine learning model that can accurately predict the likelihood of loan approval for applicants

02



To achieve a target accuracy score of 80% for the predictive model, ensuring reliable predictions that aid in decision-making.

03



To use insights from the Model to automate the loan approval system based on best features

Problem Statement

The Bank of India wishes to enhance its loan approval process by developing a predictive model that accurately determines the likelihood of loan approval for applicants.

Data Understanding

The Loan Approval Dataset was sourced from Kaggle. The dataset has comprehensive information regarding a loan applicant's financial status, personal attributes such as age and marital status

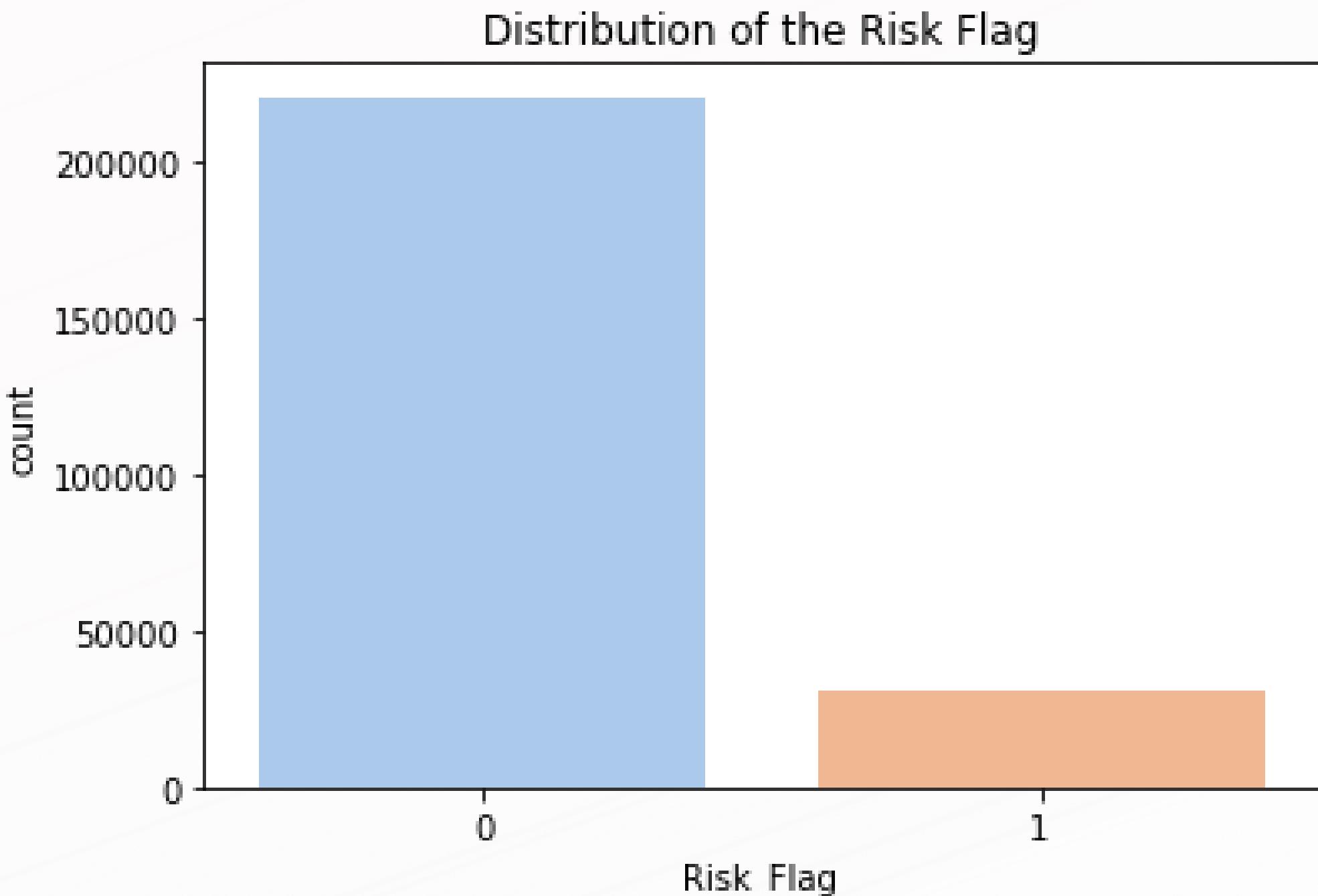


TIMELINE

The Project was Initialized on 30
May 2024 and was due on 07 June
2024

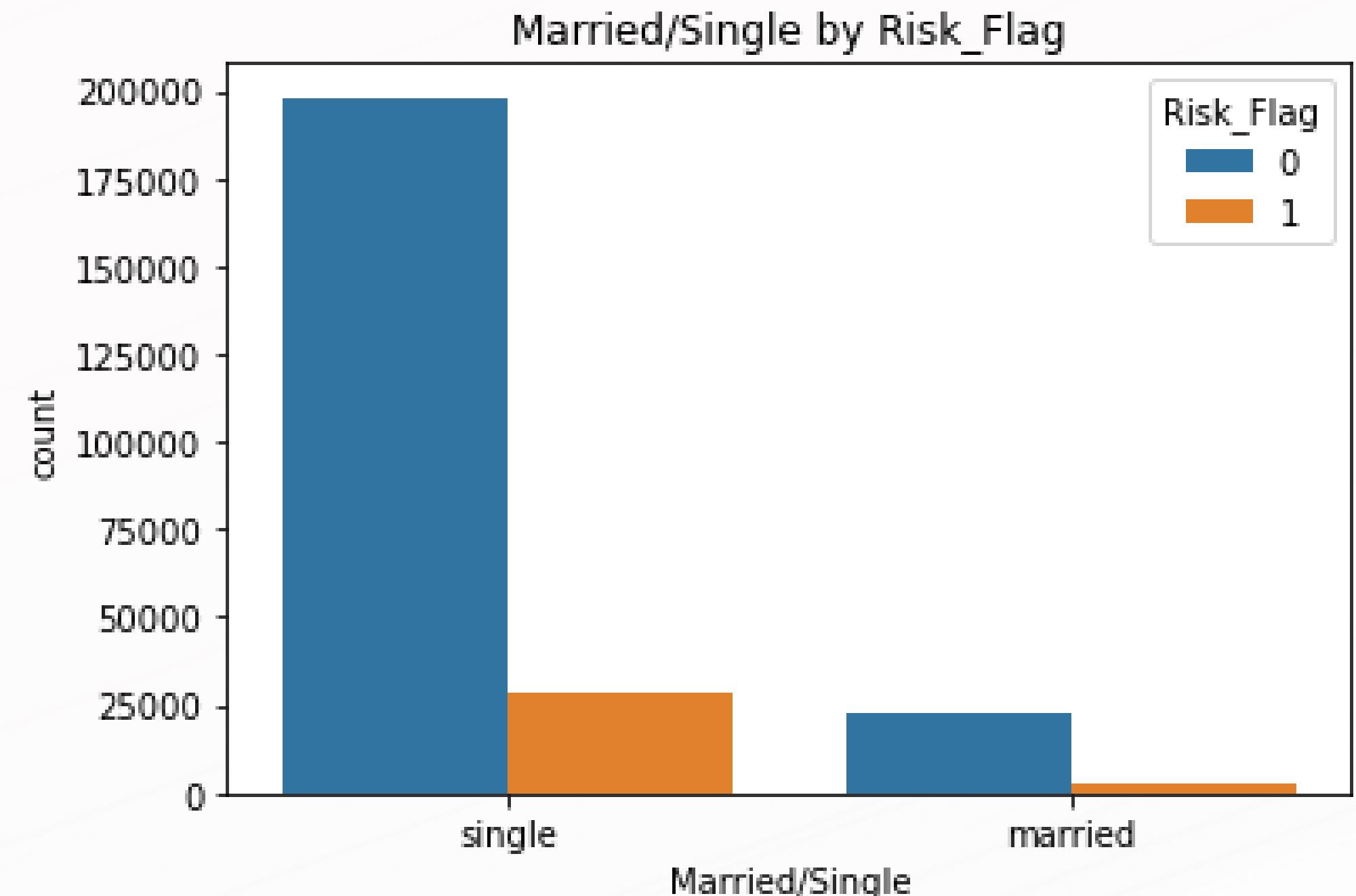
Analysis of the Risk Flag Information

More than 200,000 of the applications get approved as they are flagged non_risky



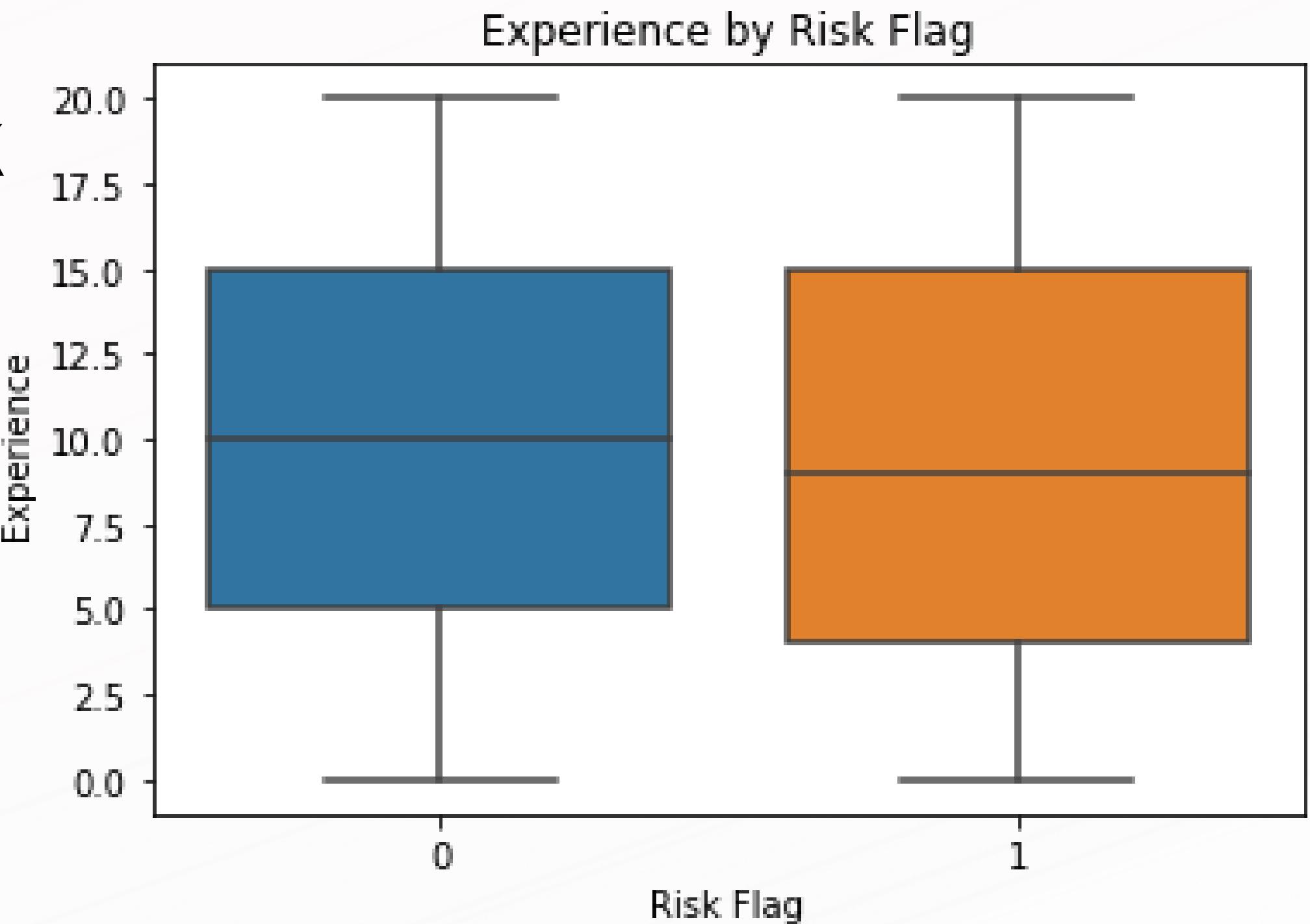
Marital Status vs Risk Flag

Most loan applicants are single and about 25,000 of the applicants are flagged as risky



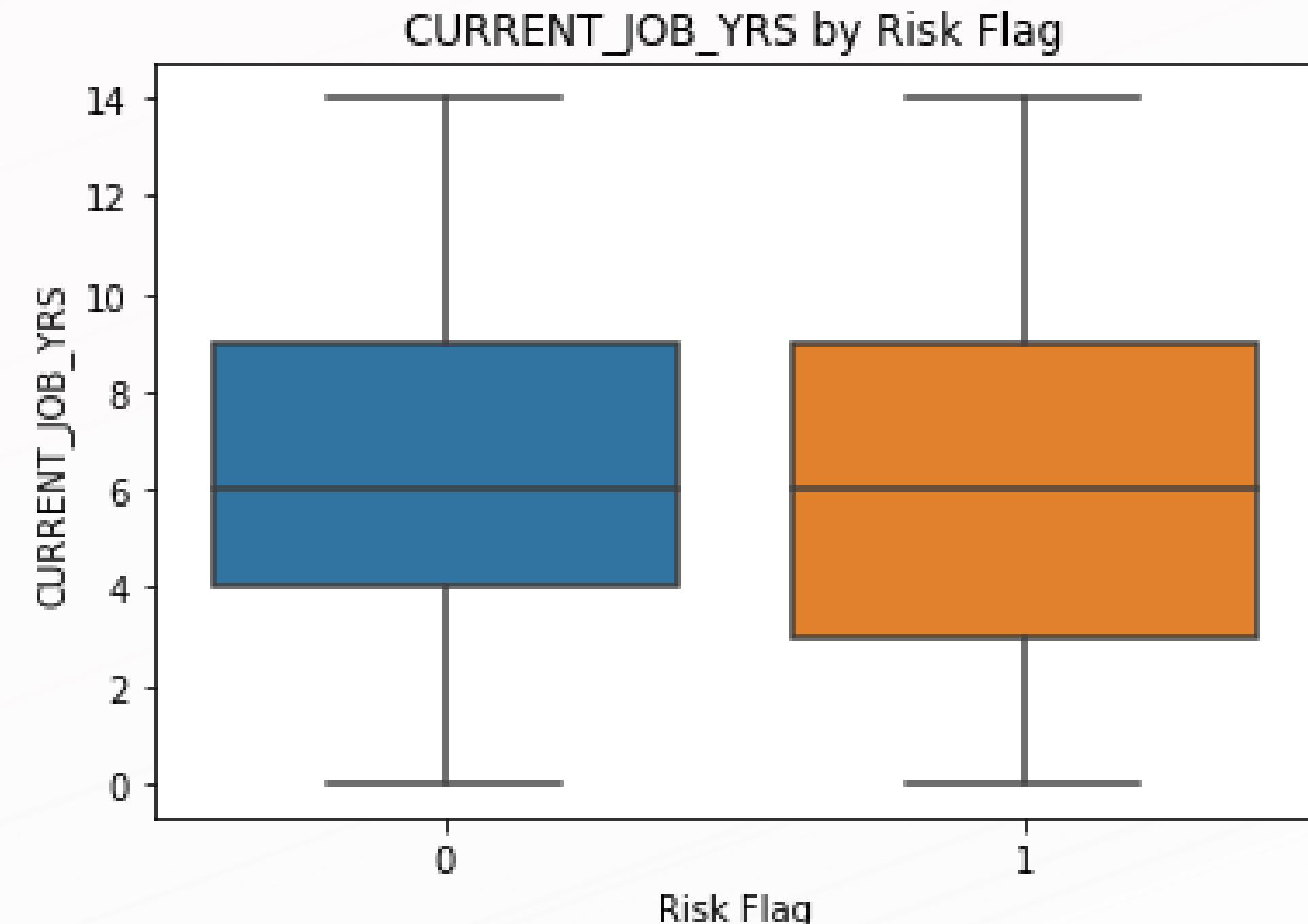
Years of Experience vs Risk Flag

Applicants with more years of experience are deemed less risky



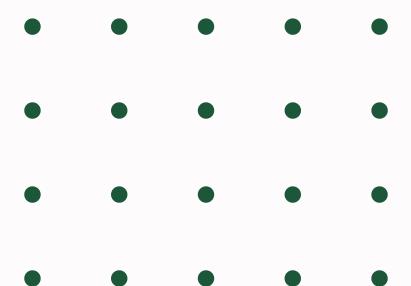
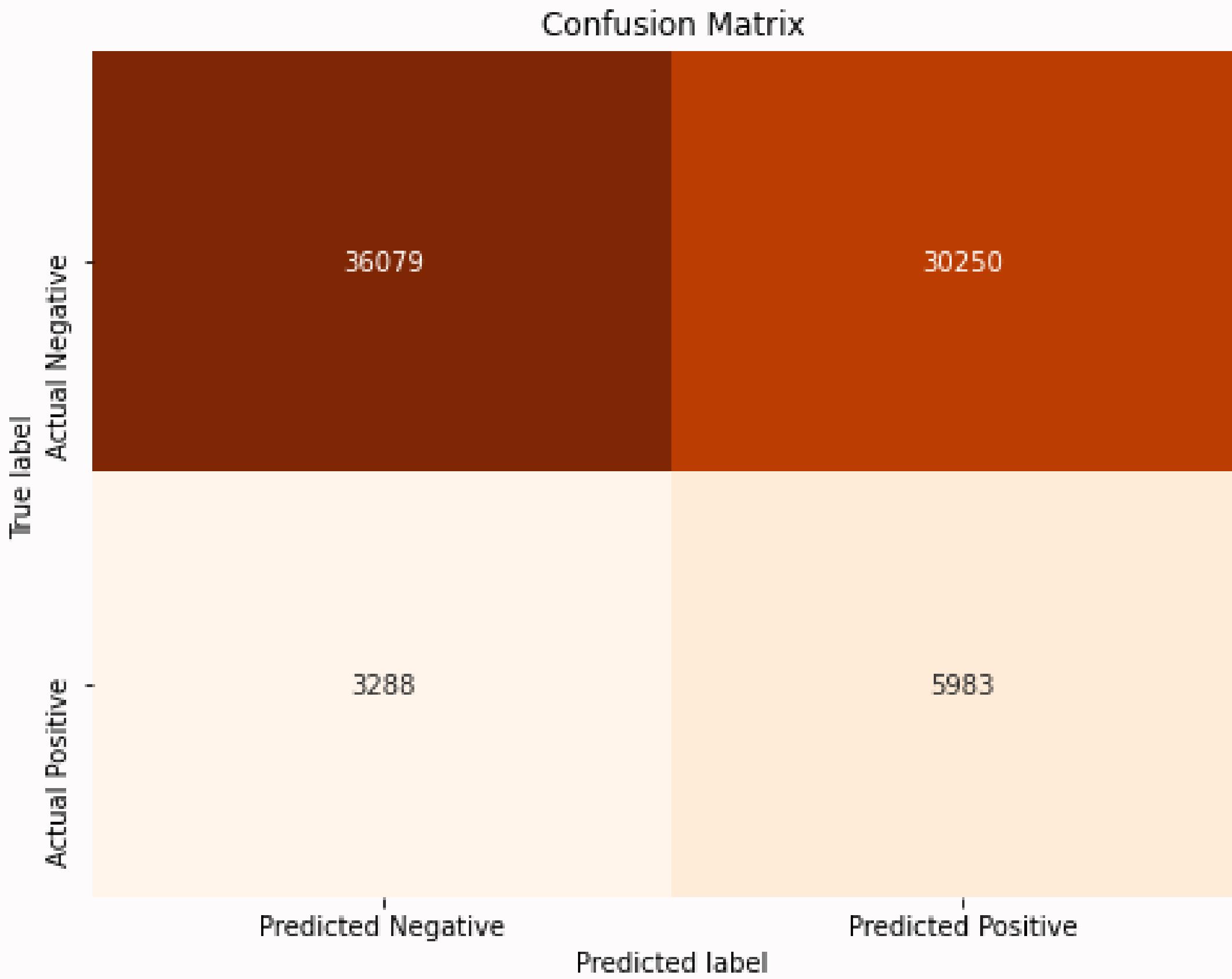
Current Job vs Risk Flag

Applicants with short job tenures are flagged as risky applicants



Logistic Regression Model

- Had an Accuracy of 55.67%
- Acted as our baseline model



Confusion Matrix

		Predicted label	Predicted Negative	Predicted Positive
True label	Actual Negative	34547	31782	
	Actual Positive	2943	6328	

2nd DecisionTree Model

- Had an Accuracy of 54.1%
- The model was not overfitting to the training dataset

Confusion Matrix

		Predicted label	
		Predicted Negative	Predicted Positive
True label	Actual Negative	45920	20409
	Actual Positive	4238	5033

Random Forest Model

Collection of Decision Trees

- Had an Accuracy of 68.9%
- The model detected 54.3% of all positive instances

Confusion Matrix

		Predicted label
		Predicted Negative
Actual label	Actual Negative	66192
	Actual Positive	137
		Predicted Positive
		7917
		1354

XGBoost Model

- Had an Accuracy of 89.3%
- For every instance it predicted a risky applicant it was correct 90% of the time

I chose this as the Best Model to Implement

Recommendations

01

Conduct fairness analysis to prevent discrimination against any group and ensure the model complies with relevant regulations

02

Provide training for stakeholders and implement a feedback mechanism to improve the model continuously

03

Continuously monitor model performance post-deployment and Plan for regular updates and retraining with new data to maintain accuracy and relevance.

04

Explore the use of advanced machine learning techniques like deep learning to further improve model accuracy and continuously explore new features that can improve the predictive power of the model.



NEXT STEPS

Introduce the company to the selected model



STEP 1

Identify and create new features that could improve the model's predictive power



STEP 2

Ensure that the model complies with relevant regulations and ethical guidelines, particularly those related to fair lending practices.



STEP 3

Set up monitoring to track the model's performance over time and implement a maintenance schedule for regular updates and retraining.



Bank Of India

THANK YOU

Hilary Simiyu

+254 712345678

hilarysimiyu@student.moringaschool.com

Moringa School 1st Floor,
Ngong Lane Plaza, Ngong Rd,
Nairobi, 00100, Kenya

