Loan Approval Classification

Welcome! Today we'll walk you through our data-driven approach to support smarter loan approval decisions.

By Brian Kanyenje

Why this Matters

Minimize Risk

Financial institutions must minimize risk. Our project uses data science. It identifies applicants likely to repay loans. This helps lenders make informed decisions.

Understanding the Business Problem

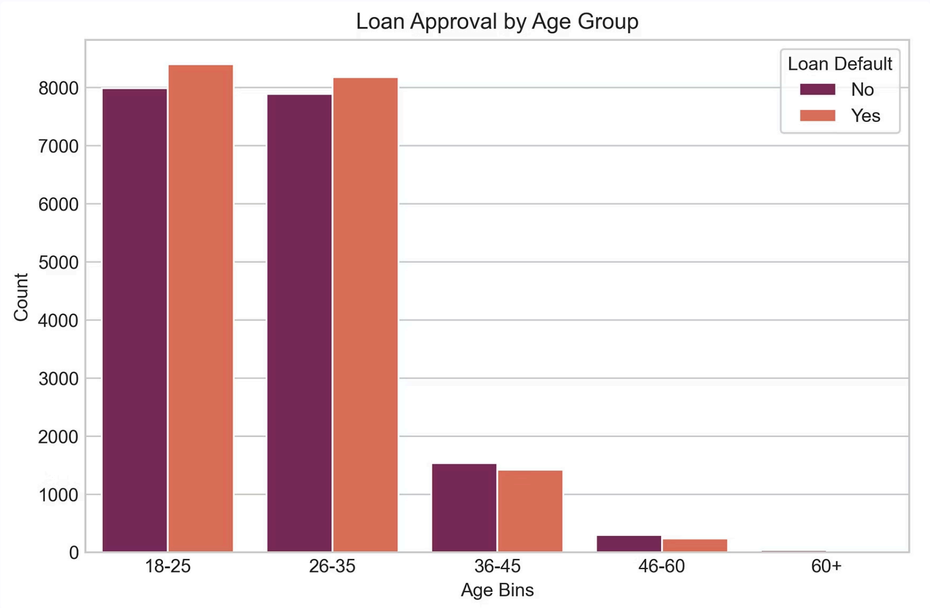
We sought to answer:

- Which demographics are more likely to default?
- Does credit score predict loan approval?
- Can we build a model to help predict loan default?

Understanding the Data

We analyzed 45,000 loan applications, each with 14 features like:

- Age, gender, education
- Income and employment experience
- Loan amount, interest rate
- Credit history and default records

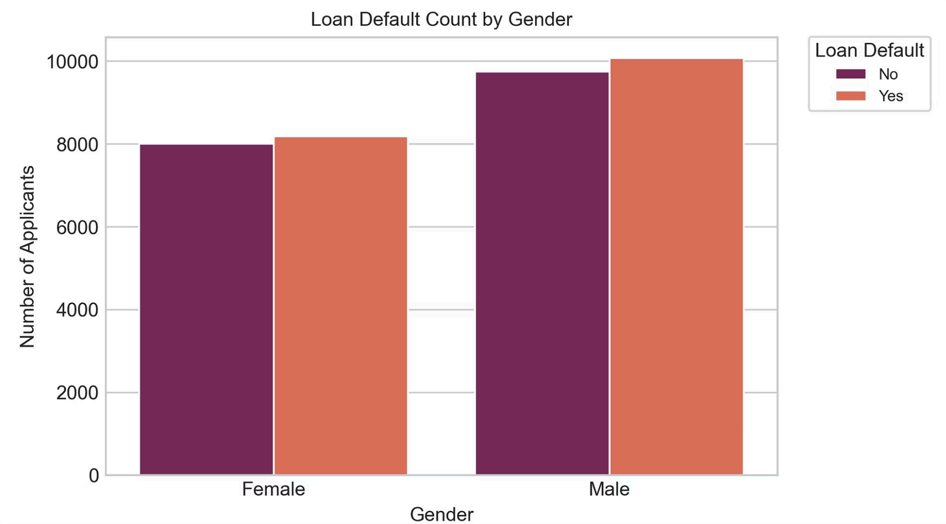


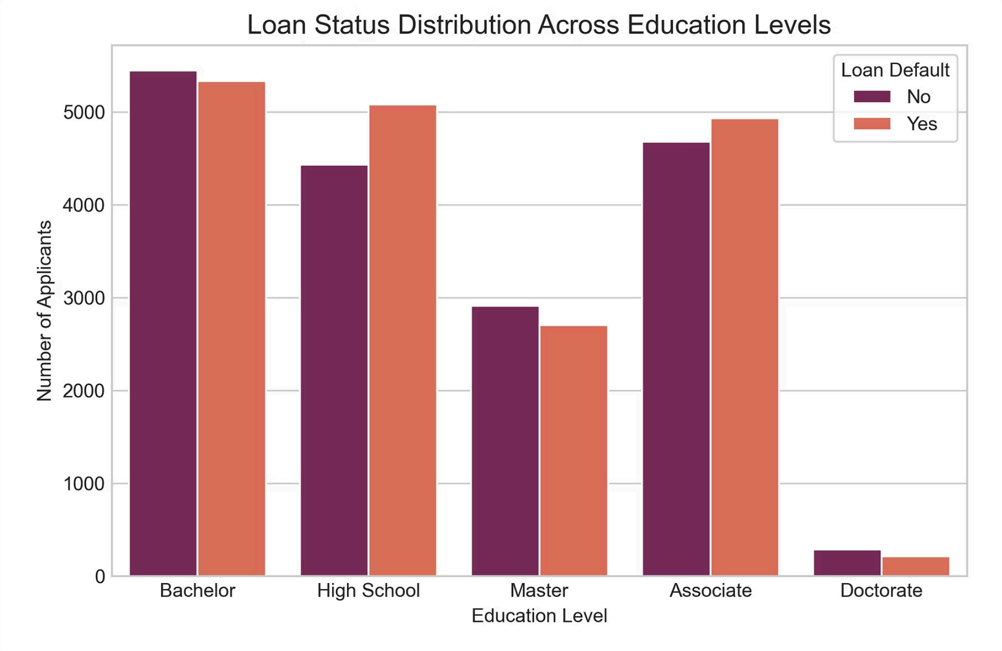
Which demographics are more likely to default?

- **Highest defaults**: Ages 18–25
- **Next highest**: Ages 26–35
- **Improved repayment**: Ages 36–45 (more non-defaulters than defaulters)
- **Sharp decline**: Loan applications drop after age 45
- **Minimal activity**: Very few loans in the 60+ age group

Which demographics are more likely to default?

- **Higher defaults among males** compared to females.
- **More male non-defaulters** as well.
- **Males dominate loan applications**, explaining higher counts in both default and non-default cases.
- **Applicants with previous defaults** show a **higher likelihood of defaulting again**.



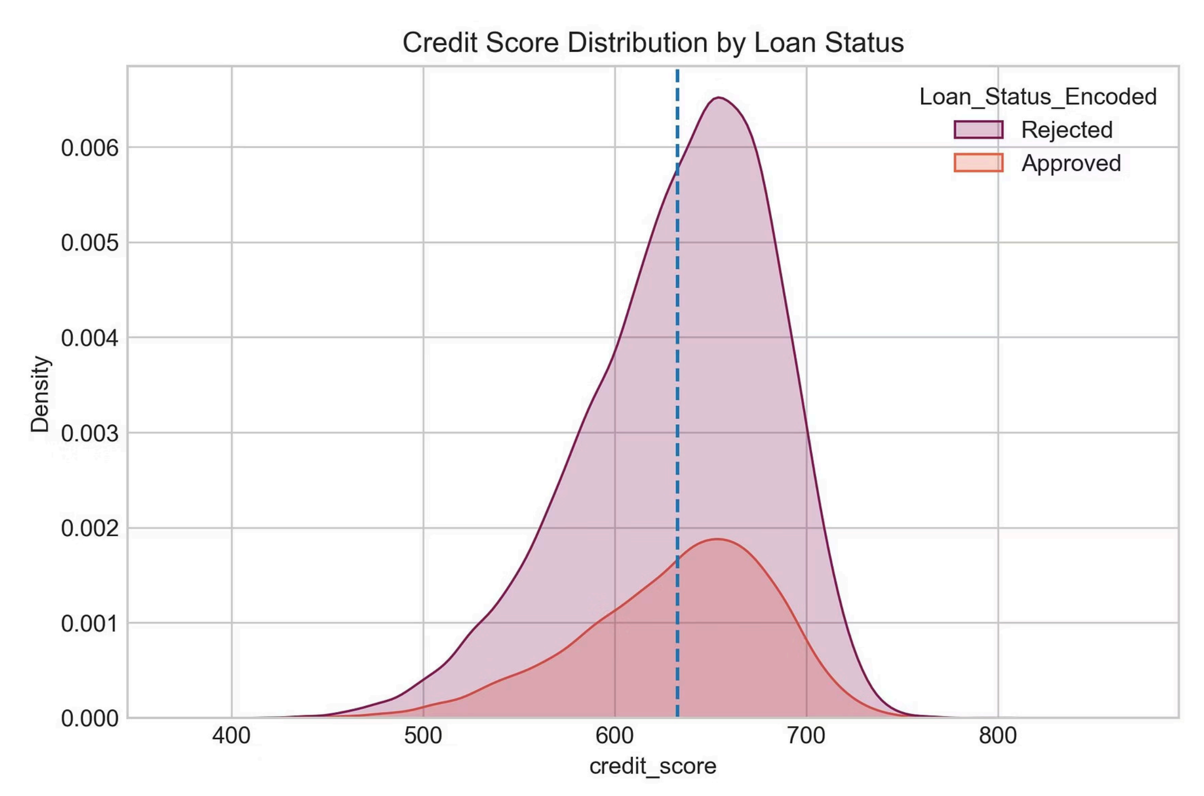


Which demographics are more likely to default?

- **Bachelor's degree holders**: Highest number of applicants; slightly more defaulters than non-defaulters.
- **High School and Associate degree holders**: More likely to default than not.
- **Doctorate holders**: Fewest applicants; very low default rates, suggesting lower risk.
- **Key insight**: Higher education doesn't always reduce default risk, except possibly at the Doctorate level.

Does credit score predict loan approval?

- **Both approved and rejected applicants** have peak credit scores around **650**.
- **Significant overlap** exists between the two groups' credit score distributions.
- **Some borrowers with 'good' scores are rejected**, while others with similar scores are approved.
- This suggests **credit score alone is not a strong predictor** of loan approval.



Building the Models

We tested three models:

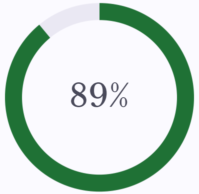
- **Logistic Regression** (baseline)
- **Decision Tree**
- **XGBoost** (advanced technique)

Each model was trained to recognize patterns between applicant features and loan outcomes.

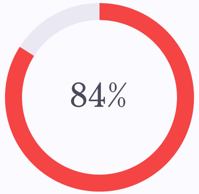
Performance Evaluation

Logistic Regression

This was the baseline model



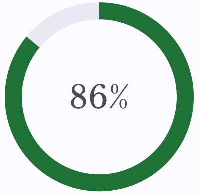
Accuracy Score



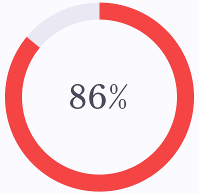
F1-Score

Logistic Regression: 89% accuracy; strong on rejections, decent on approvals.

Decision Tree



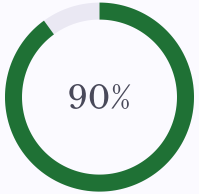
Accuracy Score



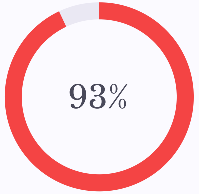
F1-Score

Decision Tree: 92% accuracy; better balance across both groups.

XGBoost - Best Model



Accuracy Score



F1-Score

XGBoost: 93% accuracy; best overall, especially at identifying approvals correctly.

What influenced Predictions

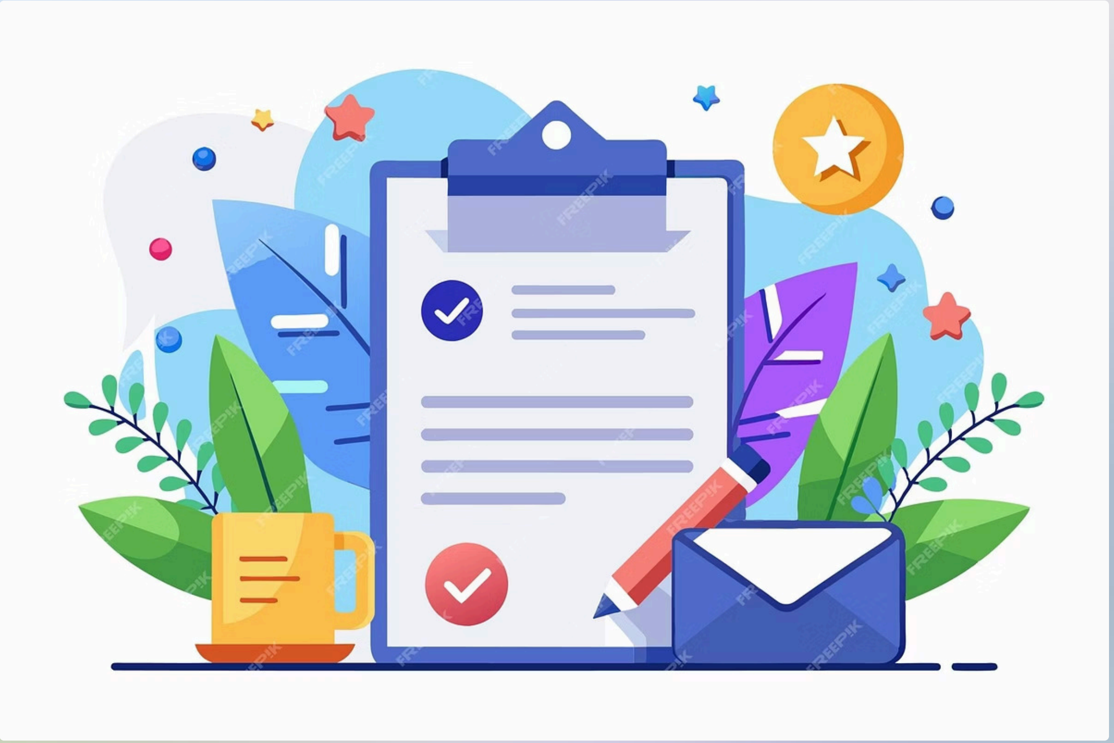
Most influential features included:

- Loan-to-income ratio
- Loan interest rate
- Employment experience
- Income level
- Previous defaults

What influenced predictions?

Most influential features included:

- Loan-to-income ratio
- Loan interest rate
- Employment experience
- Income level
- Previous defaults



Recommendations

- Consider **age** and **education level** in loan approval decisions.
- **Do not use gender** as a predictive factor.
- **Avoid relying solely on credit score** for approval decisions.
- Use the model to support decisions as it shows **high predictive accuracy**.
- The model is **robust and reliable** for real-world applications.