

Loan Prediction Analysis

Machine learning approach to predict loan approval using customer demographics and financial data



Dataset Overview

614

Total Records

Complete loan applications analyzed

13

Features

Variables including demographics, income, and credit history

4

Data Types

Mix of numerical and categorical variables



Missing Data Challenge

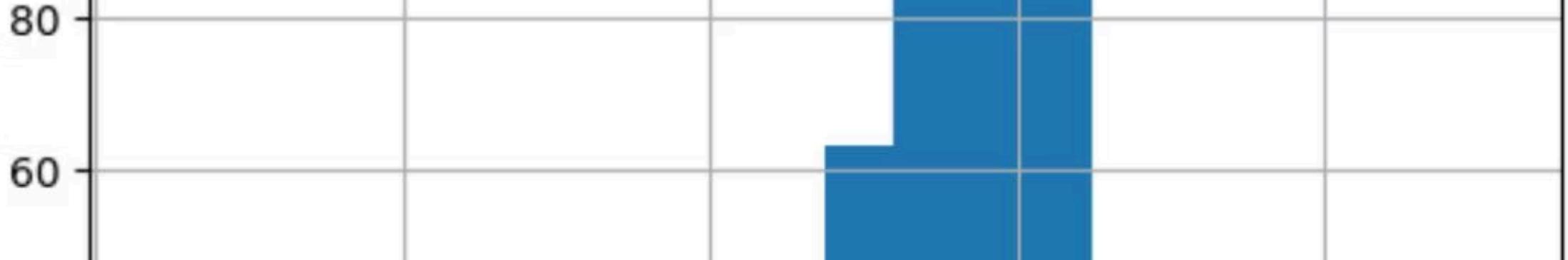
Key Missing Values

- Credit History: 50 missing
- Self Employed: 32 missing
- LoanAmount: 22 missing
- Dependents: 15 missing
- Gender: 13 missing

Solution Applied

Mode imputation for categorical variables, mean imputation for numerical features

All missing values successfully handled before model training



Feature Engineering

$$\frac{f}{dx}$$



Log Transformation

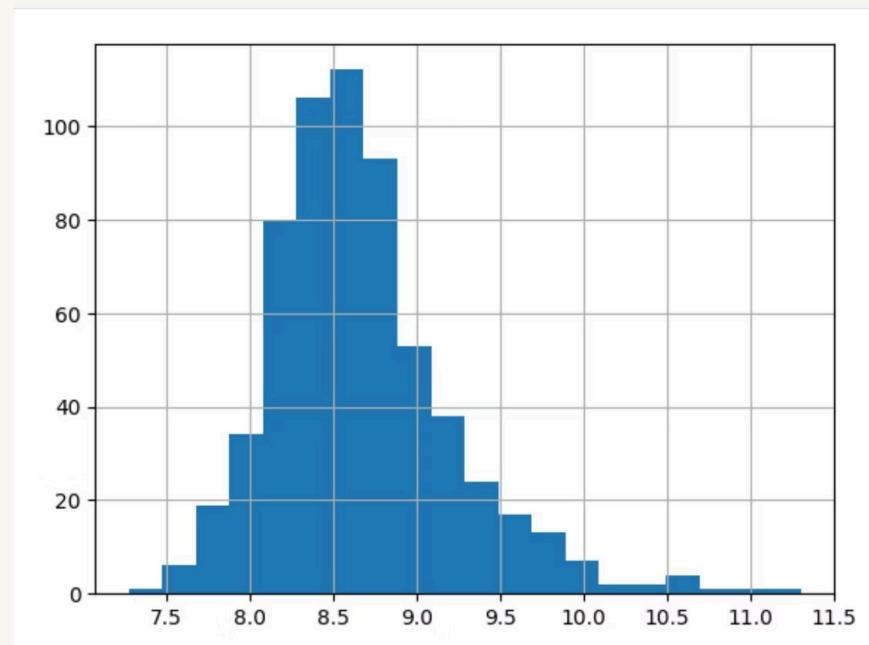
Applied to LoanAmount to normalize distribution

Total Income

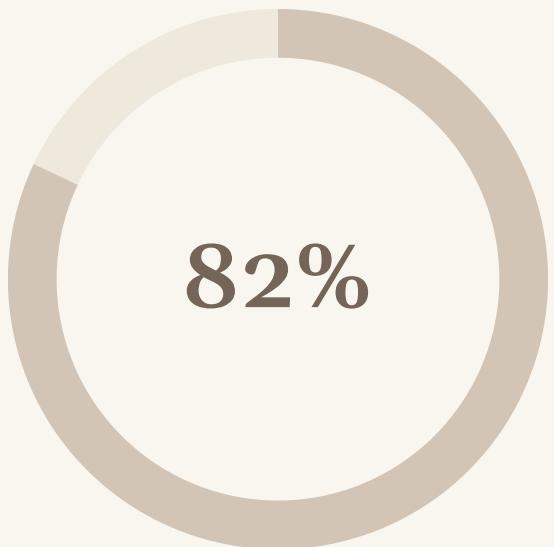
Combined Applicant + Coapplicant income

Log Total Income

Normalized total income for better model performance

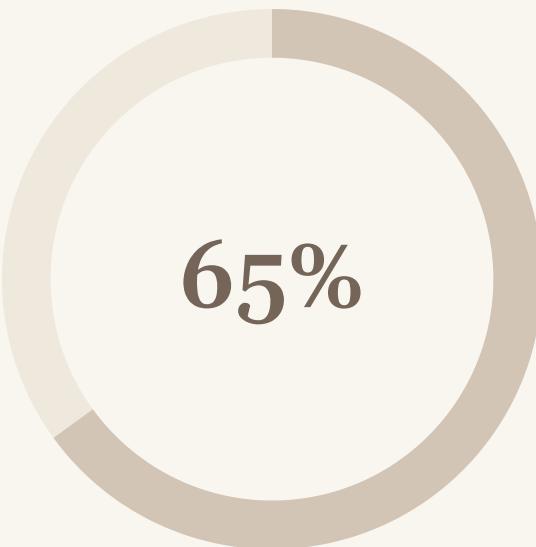


Applicant Demographics



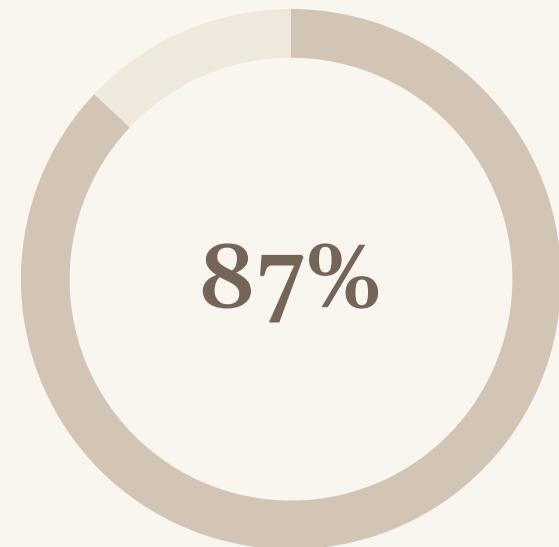
Male Applicants

502 out of 614 total



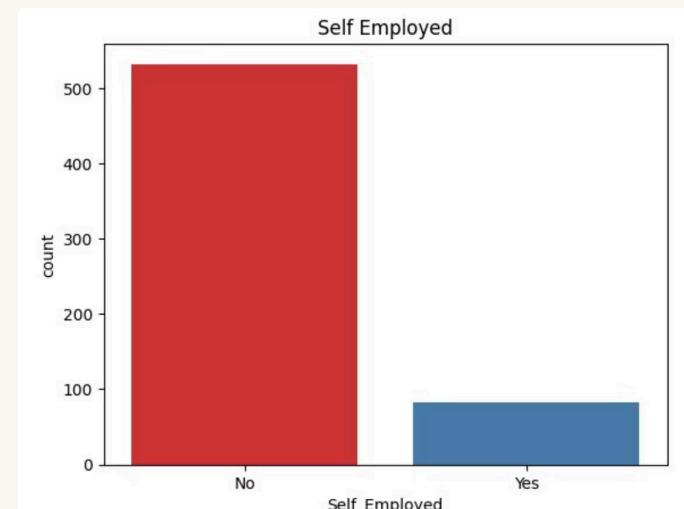
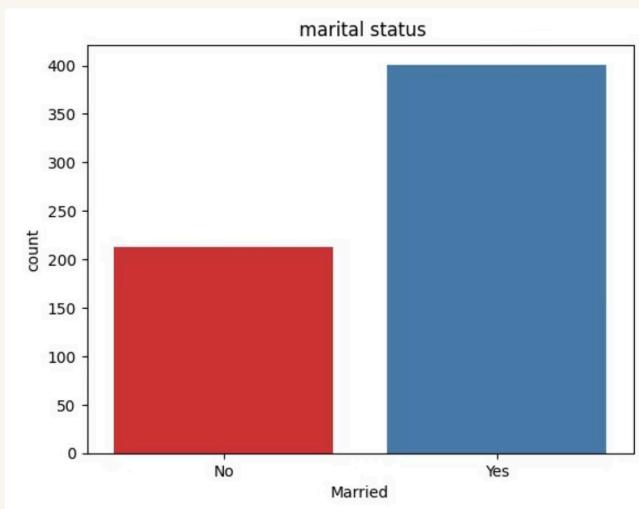
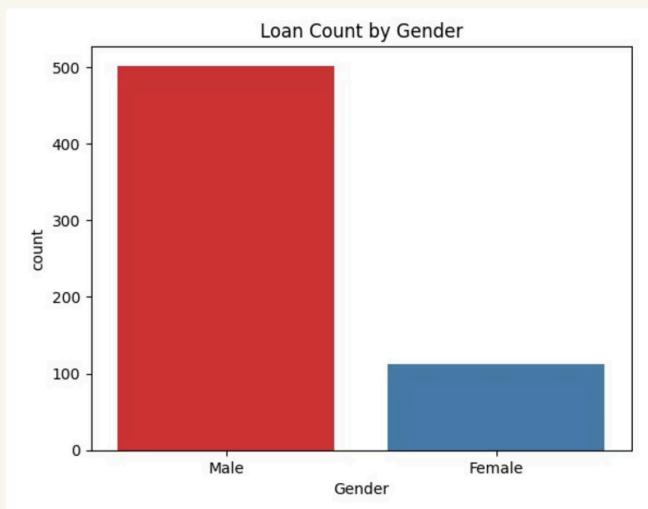
Married

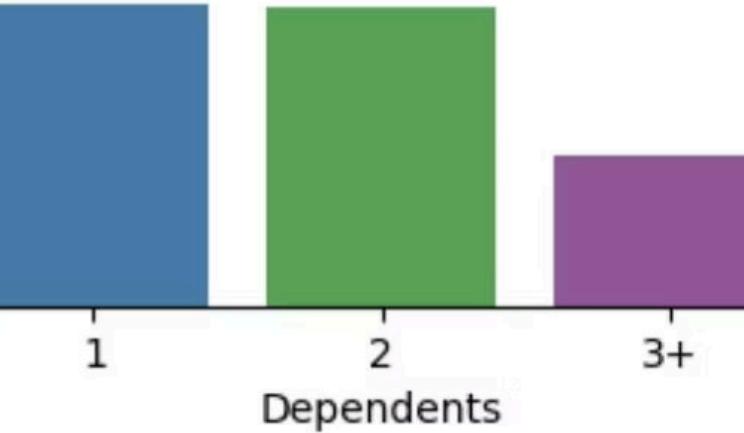
401 married applicants



Not Self-Employed

532 salaried employees





Dependents Distribution

No Dependents

345 applicants (56%)

1 Dependent

102 applicants

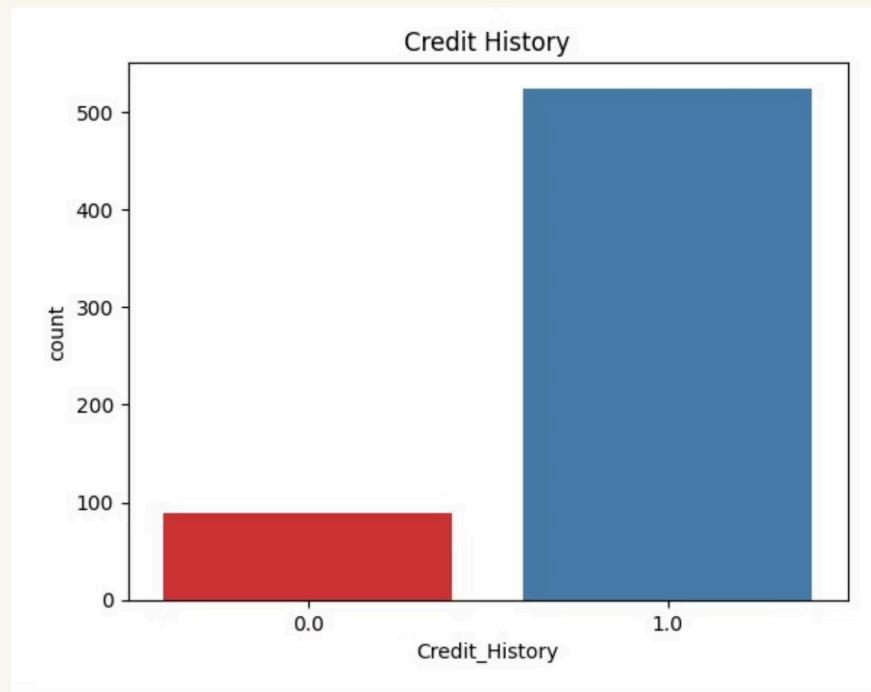
2 Dependents

101 applicants

3+ Dependents

51 applicants

Credit History Impact



Strong Predictor

525 applicants with positive credit history

Only 89 with no credit history

Critical feature for loan approval prediction

Data Preprocessing Pipeline

01

Train-Test Split

80-20 split: 491 training, 123 test samples

02

Label Encoding

Converted categorical variables to numerical format

03

Feature Scaling

StandardScaler applied to normalize feature ranges

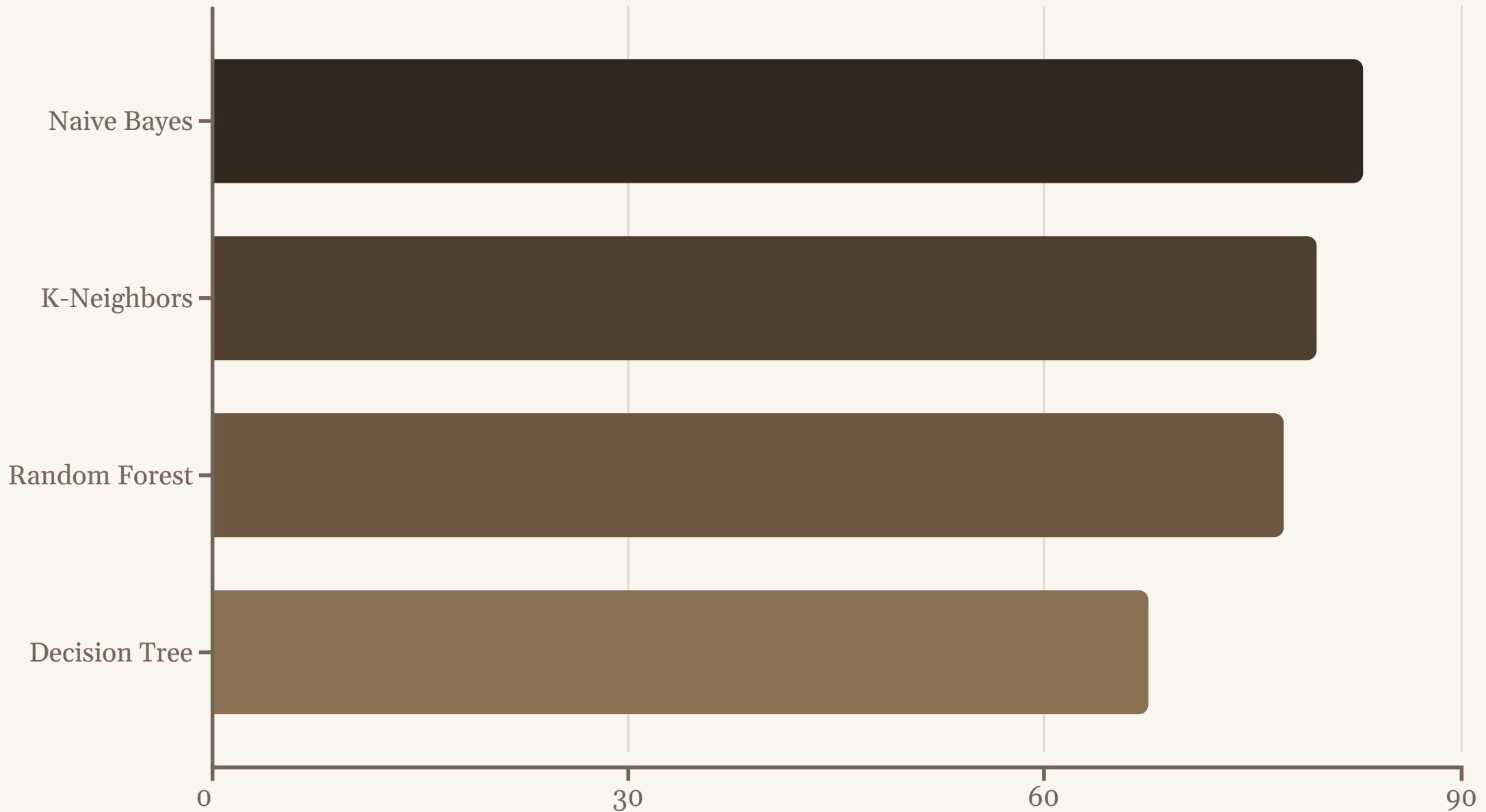
04

Model Training

Multiple algorithms tested for optimal performance



Model Performance Comparison



Naive Bayes emerged as the top performer with 82.93% accuracy, demonstrating strong predictive capability for loan approval

Key Takeaways

Best Model

Naive Bayes classifier achieved **82.93% accuracy**

Critical Features

Credit history, total income, and marital status drive predictions

Data Quality

Proper handling of missing values and feature engineering essential

Business Value

Automated loan approval screening reduces processing time and improves consistency

