Loan Approval Prediction Report

Problem Statement

Financial institutions receive thousands of loan applications every day. Manually evaluating each is time-consuming and error-prone. This project aims to automate loan approval decisions using historical loan data and machine learning.

Objective

- Predict loan approvals based on applicant data.

- Discover key data patterns affecting approval decisions.

- Build a clean, production-ready ML pipeline for real-world deployment.

Analysis & Engineering Steps

Step	Objective
Load & Inspect Data	Understand data structure and missing values
Clean Missing Values	Impute nulls using mode/median logic
Feature Engineering	Create new variables like Total_Income, EMI
Exploratory Analysis	Discover trends in gender, income, credit history
Outlier Treatment	Cap extreme values to improve robustness
Model Building	Train a Logistic Regression classifier
Model Evaluation	Use accuracy, precision, recall, F1-score
Model Comparison	Evaluate against Random Forest, SVM, Naive Bayes

Model Results (Logistic Regression)

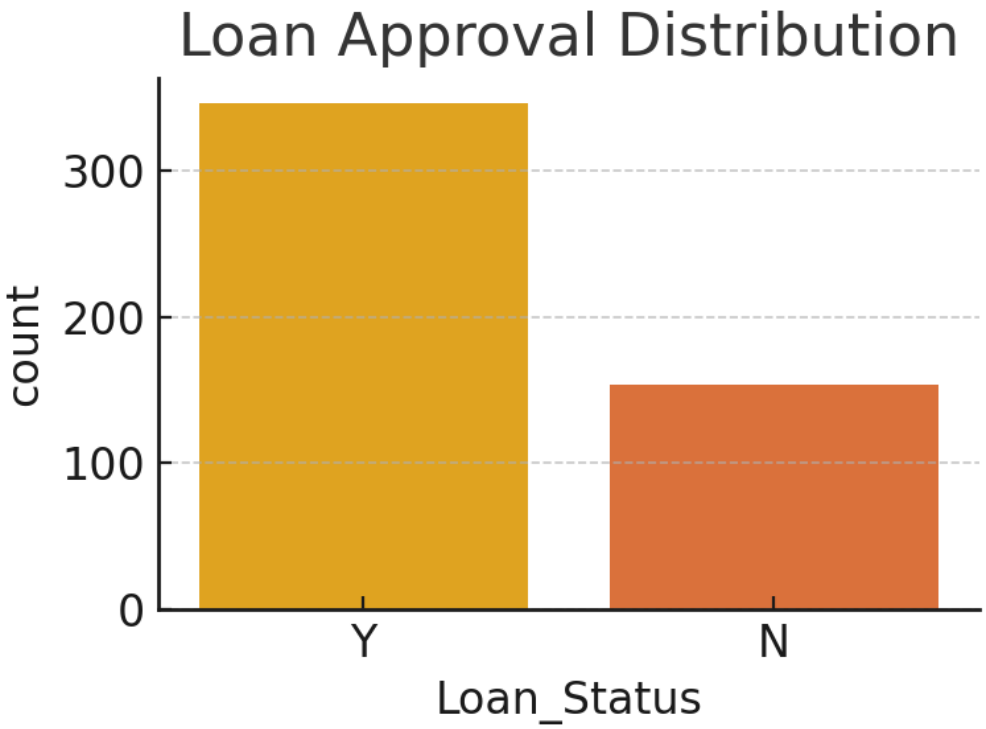
- Accuracy: 79.67%
- Precision: 76.19%
- Recall: 100%
- F1-Score: 86.49%
- Chosen Model: Logistic Regression

Business Insights

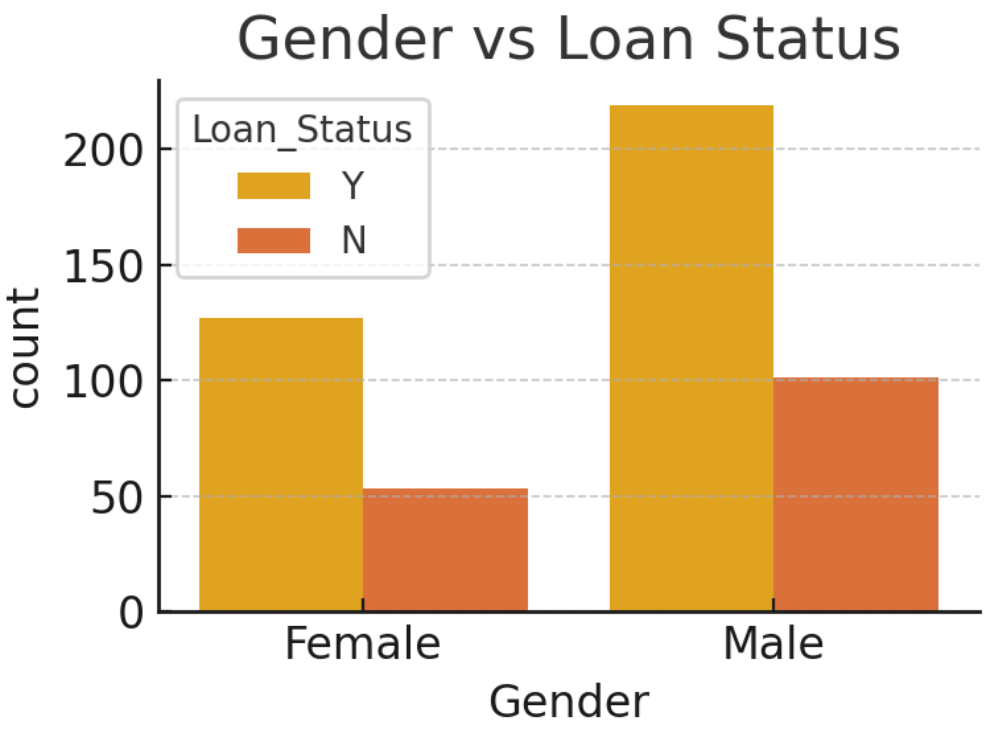
- Credit History is the strongest predictor.
- High income-to-EMI ratio improves approval chances.
- Semiurban property areas have the highest approval rates.
- Self-employed applicants face more scrutiny.
- High loan amounts relative to income are riskier.

Key Visualizations

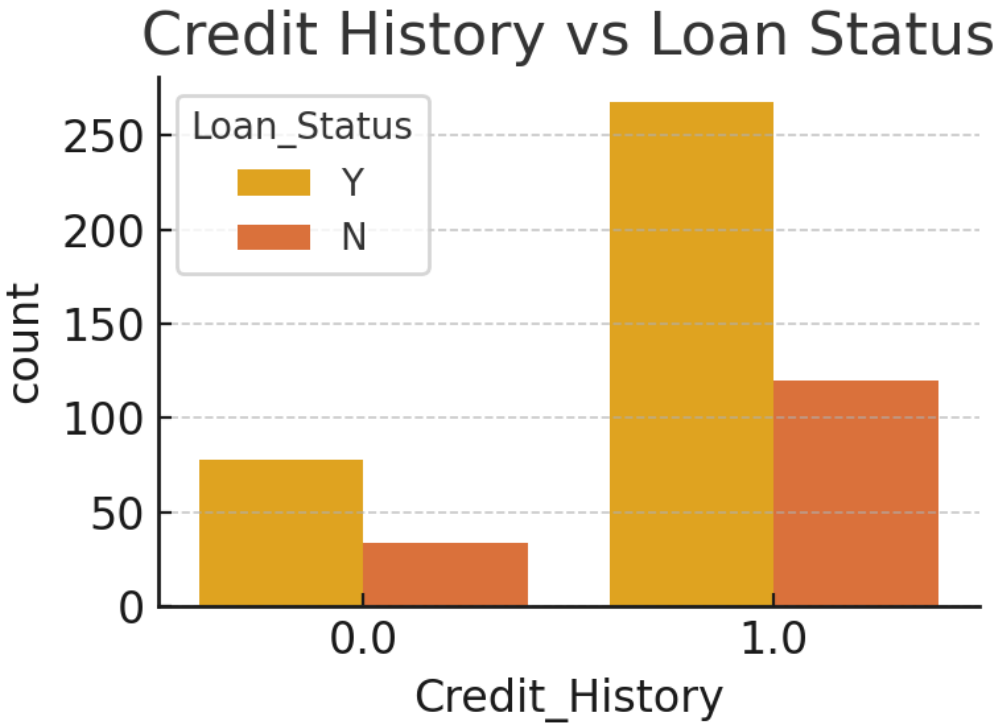
Loan Approval Distribution



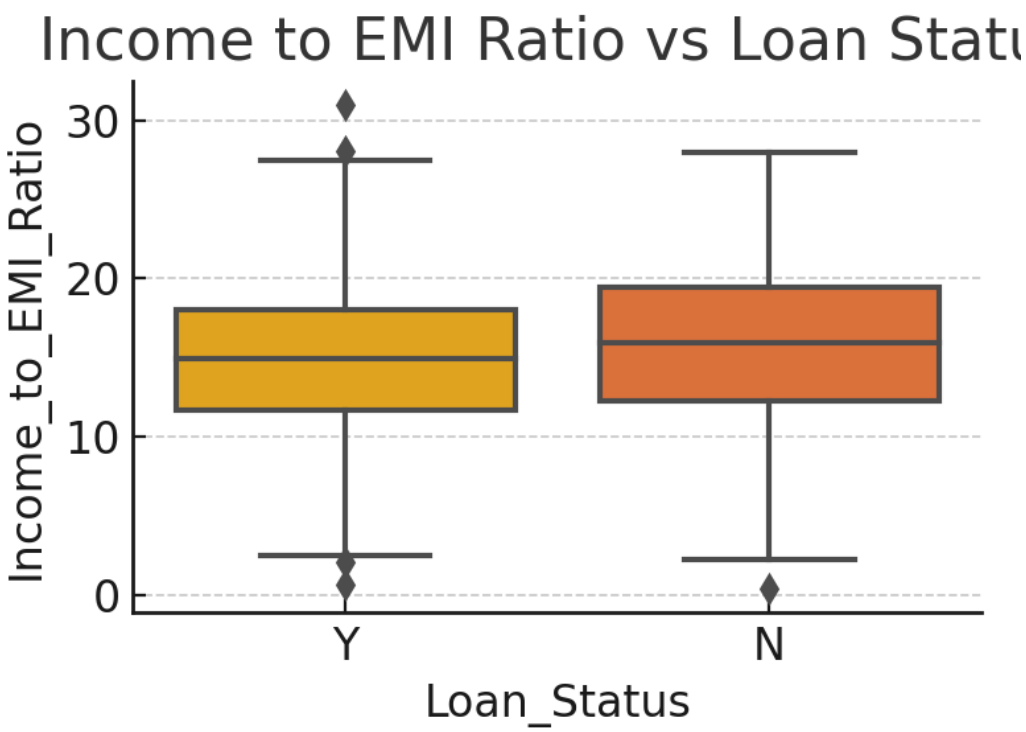
Gender vs Loan Status



Credit History vs Loan Status



Income to EMI Ratio vs Loan Status



Why This Project Stands Out

- End-to-End Workflow: raw data to cleaned dataset to EDA to model to insights.

- Production-Friendly Code: modular, reusable, and scalable.

- Data Engineering Skills: pipeline thinking, version control, reproducibility.

- Real-World Use Case: simulates how banks assess risk and automate approvals

About Me

I'm actively transitioning into a Data Scientist role, with a deep passion for Cloud technologies and AI tools which I am learning alongside my current projects.

This project showcases my ability to:
- Build end-to-end machine learning pipelines
- Analyze business problems and drive actionable insights
- Engineer scalable and interpretable models for real-world deployment

Email: harshkumar7649@gmail.com
LinkedIn: https://www.linkedin.com/in/harsh-b50377206/
GitHub: https://github.com/harshnu2020