Loan Default Prediction Project Report

# 1. Introduction

This report presents a data science project focused on predicting whether a loan applicant will default or repay the loan. Using historical data and machine learning, the model helps financial institutions make informed lending decisions.

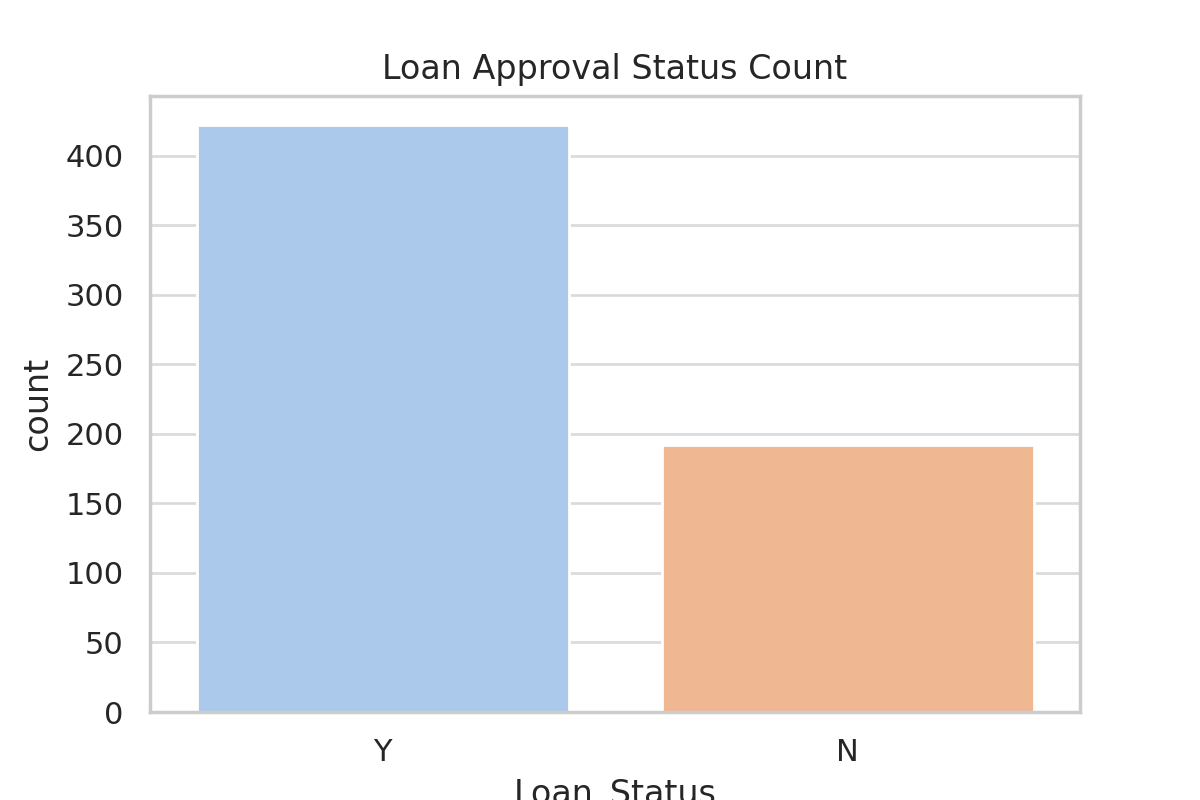
# 2. Dataset Overview

The dataset consists of demographic and financial information of loan applicants. Key columns include:  
- Gender, Marital Status, Dependents  
- Education, Self-Employed  
- ApplicantIncome, CoapplicantIncome  
- LoanAmount, Loan\_Amount\_Term  
- Credit\_History, Property\_Area  
- Loan\_Status (Target Variable)

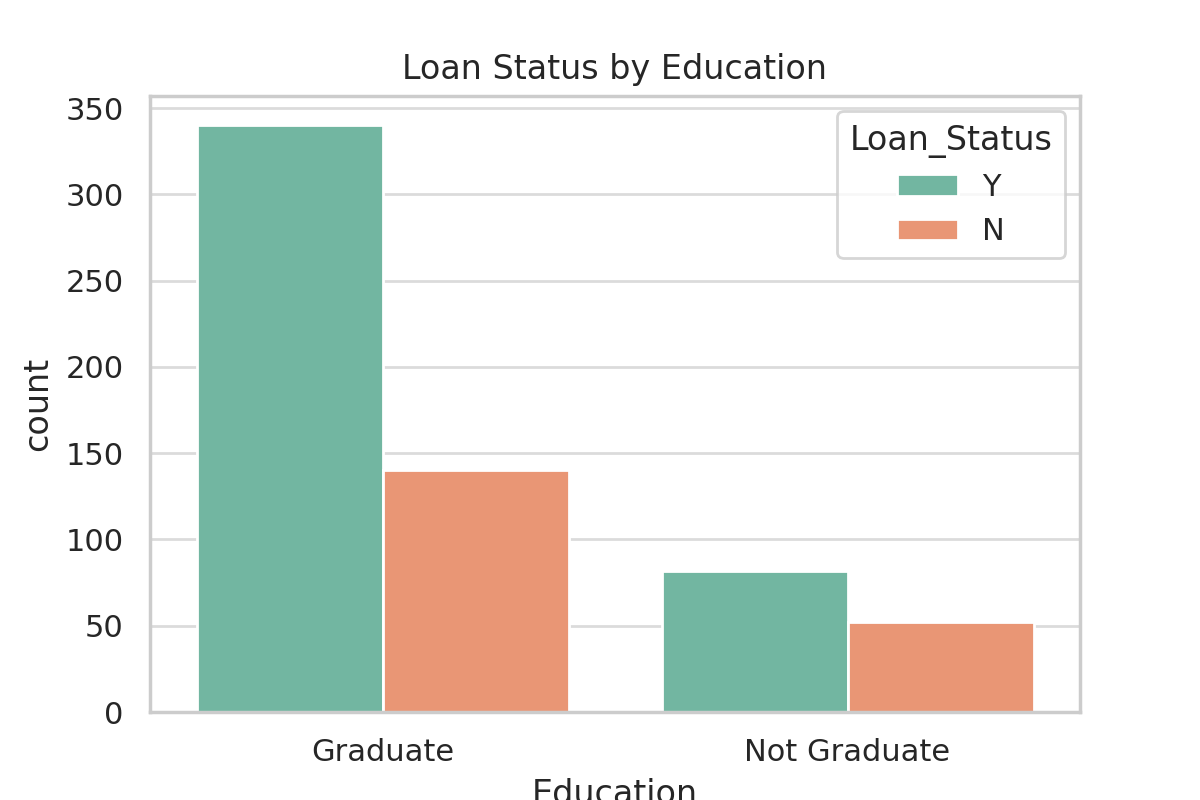
# 3. Exploratory Data Analysis (EDA)

Initial analysis was conducted to understand trends and relationships between the variables. Key findings are shown below:

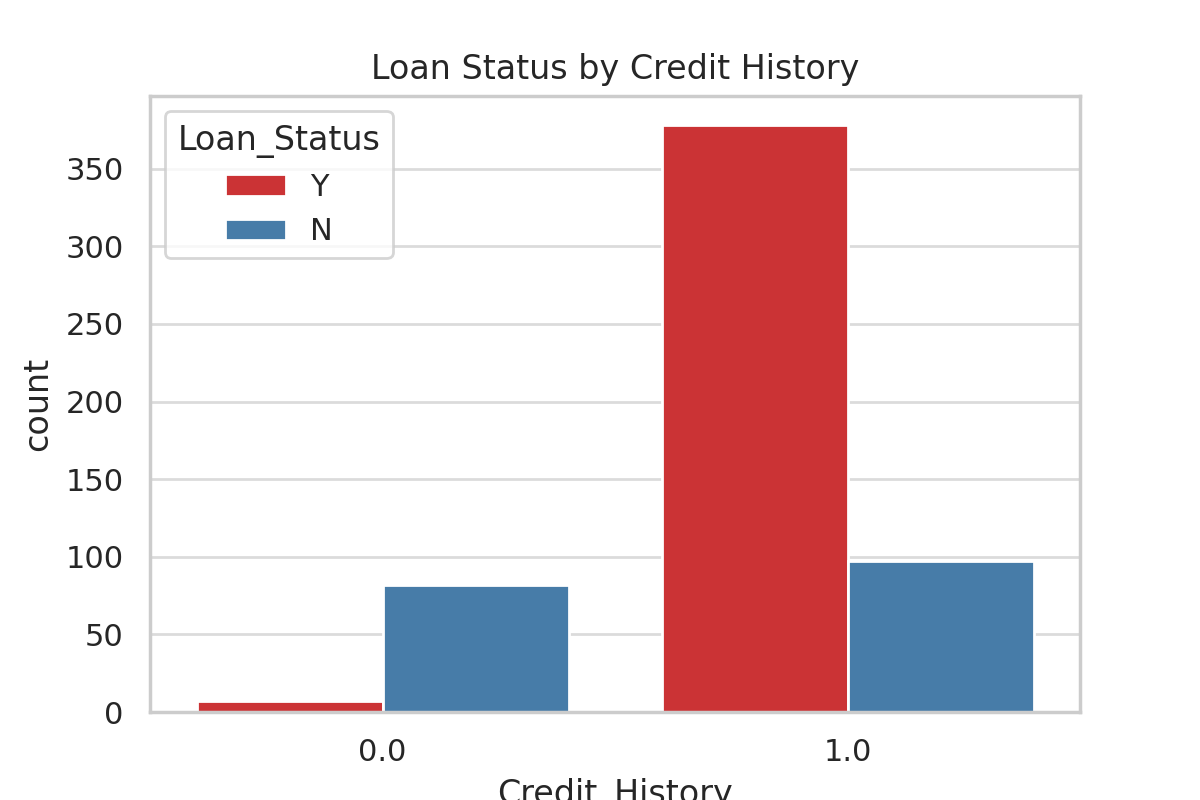
Loan Approval Status Count:



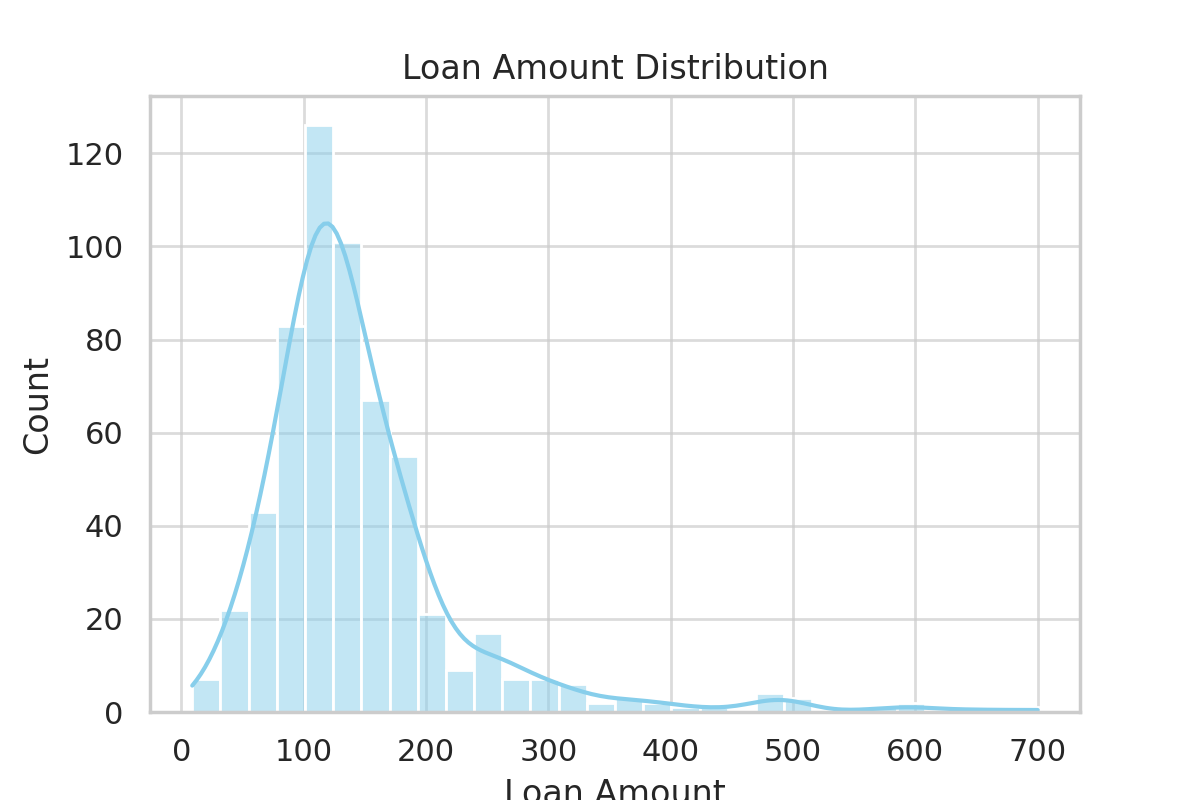
Loan Status by Education:



Loan Status by Credit History:



Loan Amount Distribution:



# 4. Machine Learning Modeling

Several models were evaluated to predict loan default status. The main model used is Logistic Regression due to its strong performance and interpretability.  
  
Model Trained:  
- Logistic Regression  
- Decision Tree  
- Random Forest  
  
Best Performing Model: Logistic Regression  
- Accuracy: ~81%  
- Important Features: Credit\_History, ApplicantIncome, LoanAmount, Education

# 5. Deployment

The final model was deployed using Streamlit, a Python framework that creates interactive web applications. Users can input applicant details and receive a prediction instantly.

# 6. Conclusion

This project demonstrates a complete machine learning workflow from data exploration and cleaning to modeling and deployment. The predictive model can support financial institutions in making smarter, data-driven loan decisions and minimizing default risks.