Predicting Loan Approvals

Ali Bahrami

Dec. 10, 2021

Agenda

- Project Overview
- Exploratory Data Analysis
- Models and Evaluation
- Conclusion & Results

01

Project Overview

Why Automate Loan Approvals?

1. Eliminate unnecessary paperwork throughout the lending process
2. Get rid of any human error in the entire crediting process
3. Reduce operational cost of running a lending business
4. Reduce time-to-market and related-costs
5. Increase Customer Satisfaction

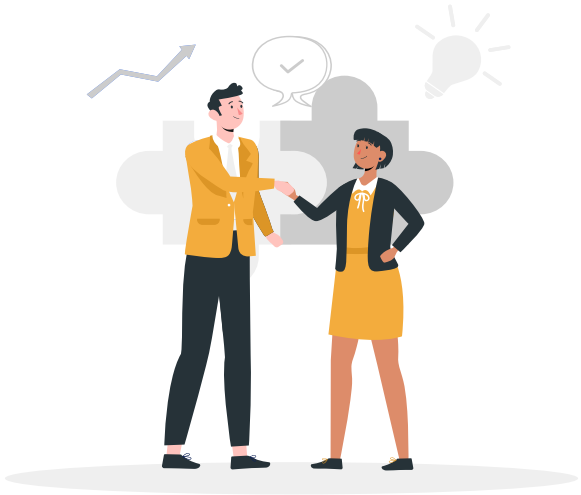
02

Exploratory Data Analysis

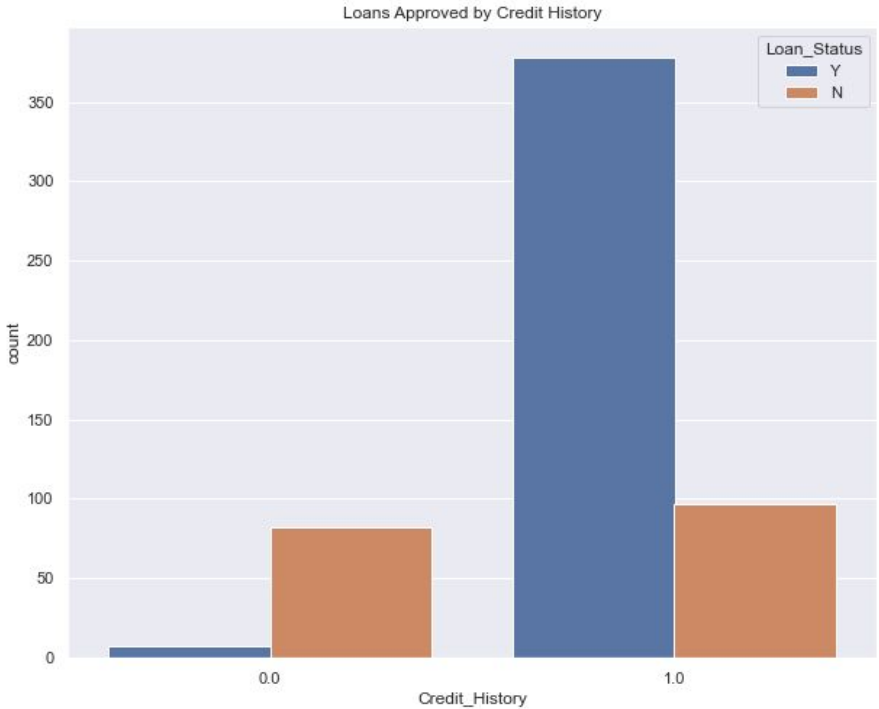
Exploratory Data Analysis

Hypothesis: Factors Influencing Loan Approvals

1. Credit history
2. Higher applicant and co-applicant incomes
3. Higher education levels
4. Properties in urban areas with high growth perspectives

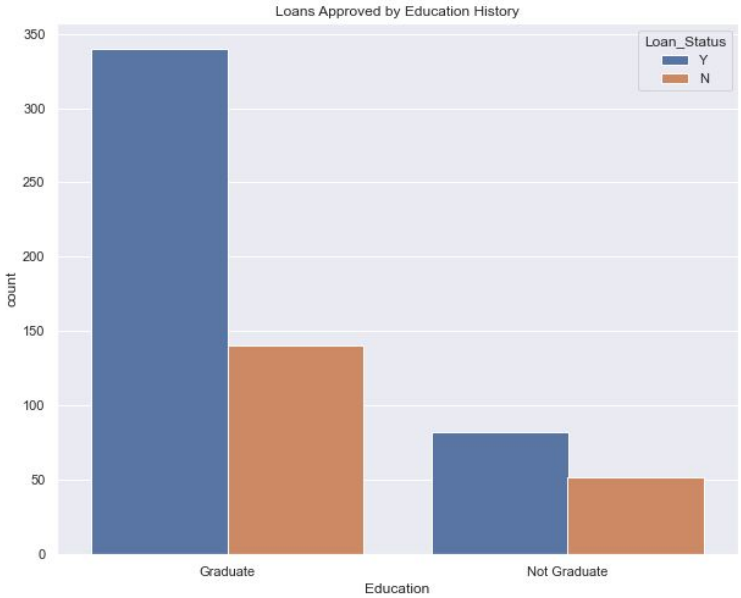


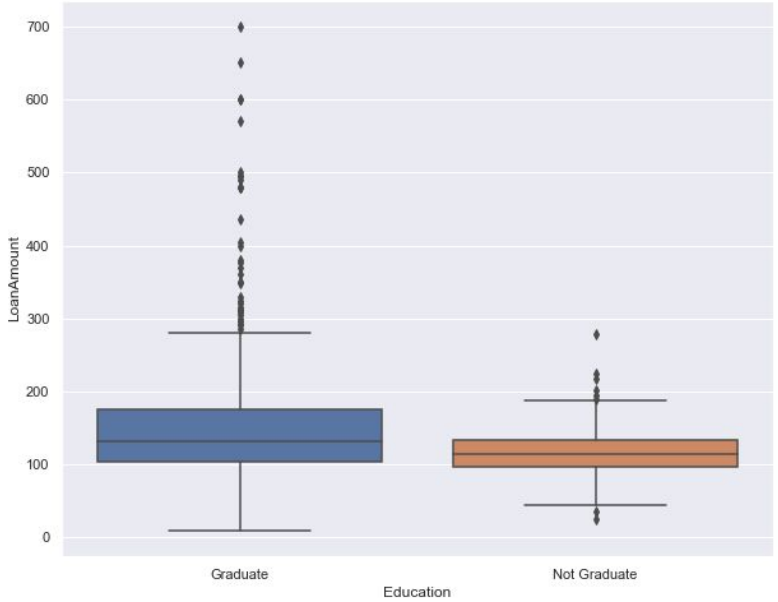
Loan Approvals by Credit History



Applicants with credit history are way more likely to get approved.

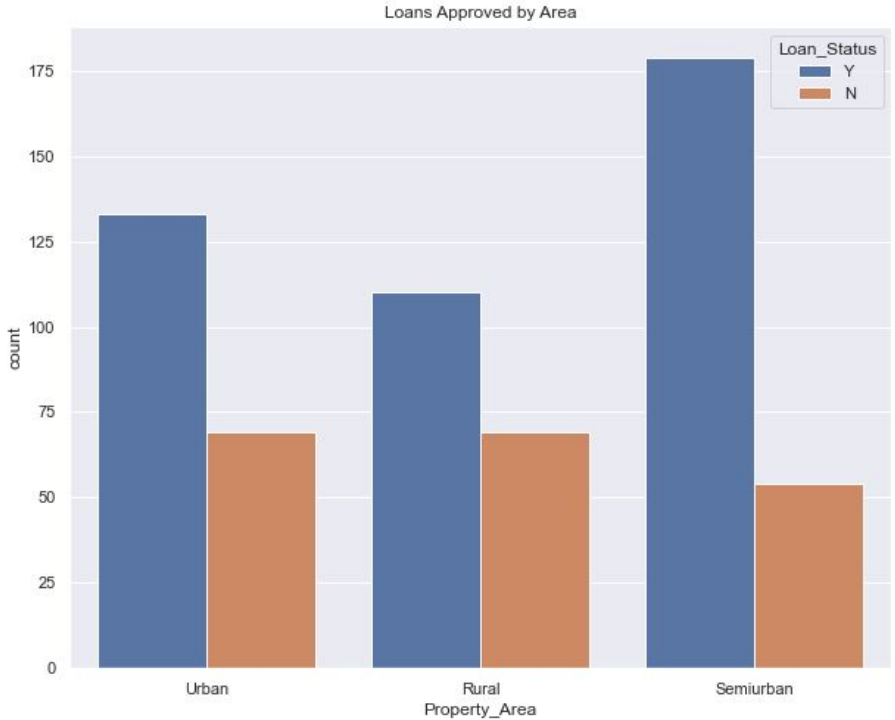
Education Stats





Applicants with higher education levels get approved more often and receive higher amounts of loan.

Loan Approvals by Area



Applicants in suburban areas have the highest loan approval rates.

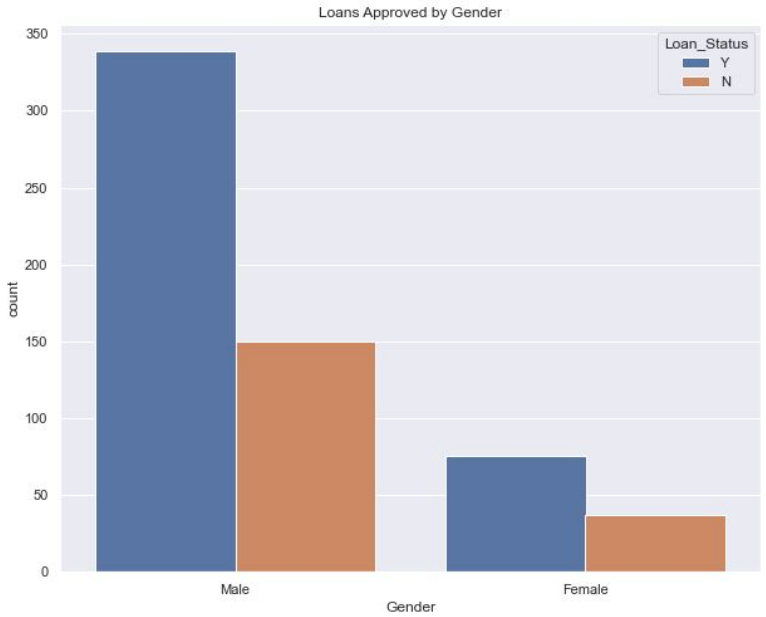
Gender Stats

Average Income

Gender	Female	Male
Education		
Graduate	3638.5	4166.0
Not Graduate	3208.5	3357.0

Average Loan Amounts

Gender	Female	Male
Education		
Graduate	113.0	136.0
Not Graduate	102.0	115.0

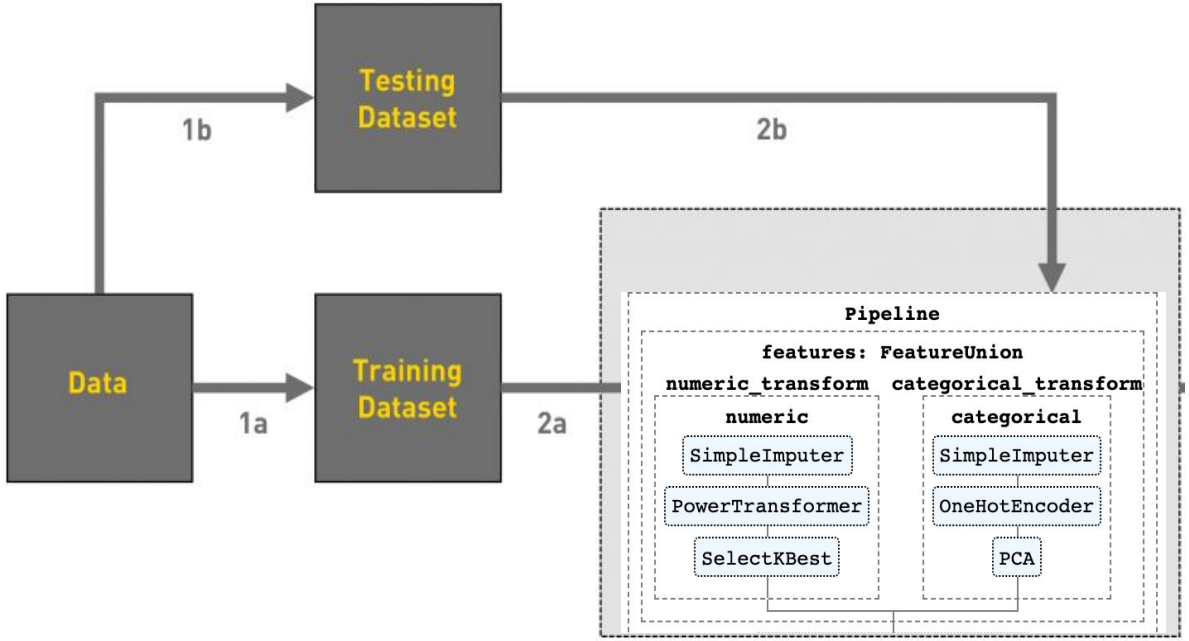


Female applicants have lower incomes, are less likely to get approved, and receive lowers amounts of loans compared to Male applicants.

03

Model & Evaluation

Pipeline Development



1. **RandomForestClassifier**
2. LogisticRegression
3. GradientBoostingClassifier
4. KNeighborsClassifier
5. CategoricalNB

04

Conclusion & Results

Results

Best Model

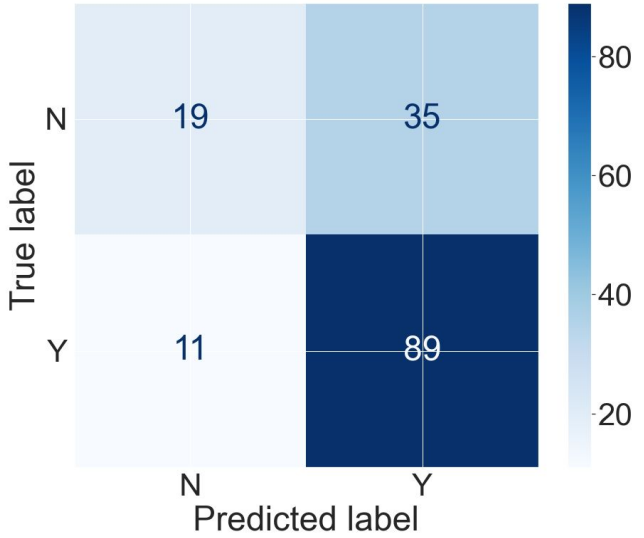
Random Forest Classifier

Train Accuracy Score

74.9%

Test Accuracy Score

70.13%



Test Case

Applicant


```
"Loan_ID":"LP002197",
"Gender":"Male",
"Married":"Yes",
"Dependents":"2",
"Education":"Graduate",
"Self_Employed":"No",
"ApplicantIncome":5185,
"CoapplicantIncome":0.0,
"LoanAmount":155.0,
"Loan_Amount_Term":360.0,
"Credit_History":1.0,
"Property_Area":"Semiurban",
"Loan_Status":"Y"
```


```
import requests
URL = "http://10.0.0.183:5005/predict"
# sending get request and saving the response
r = requests.post(url = URL, json = json_data)
```

✓ 0.1s

```
print(r.json())
```

✓ 0.1s

```
{'Loan Approved: ': "['Y']"}
```

Going Forward

1. Imputing missing values separately for each variable
2. Incorporating other scaling techniques in GridSearch (time-consuming)
3. Deployment to the cloud with a user-friendly interface

Thank you!

Presentation Template From: SlidesGo