



Loan Approval Predictions Using Logistic Regression

Isaac Turk

Background



As of September 30, 2024, Synchrony reported total loan receivables of **\$102.193 billion**



Consumer installment loans and credit card receivables each accounted for **\$10.286** billion of the total loan receivables



Synchrony's net charge-off rate for the full year 2023 was **4.84%**, which was below their target underwriting range of **5.5%-6.0%**



My goal with this mini-project is to help lower that charge-off rate even more by creating a model to classify if a customer is likely to default

About the Data



Dataset contained **45,000** **observations** and **14 variables**



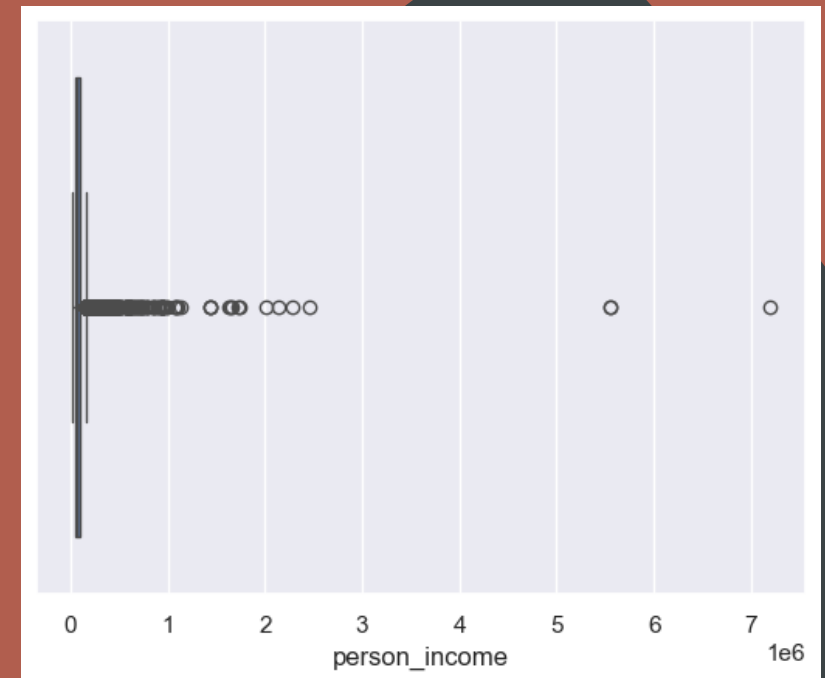
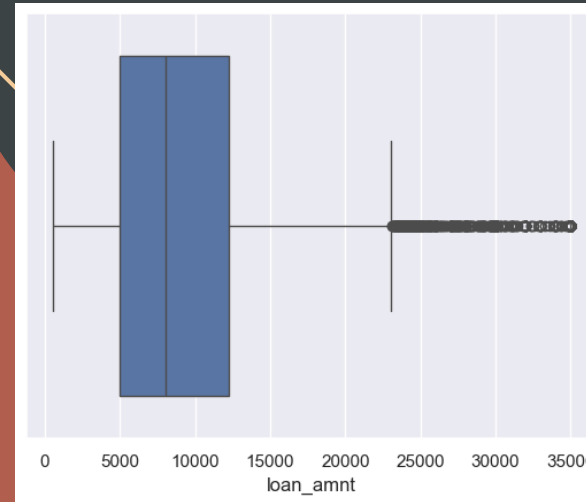
Synthetic dataset created to understand financial risk factors and simulating loan approval models



Each observation is a loan request (age, education, credit history, income, etc.)

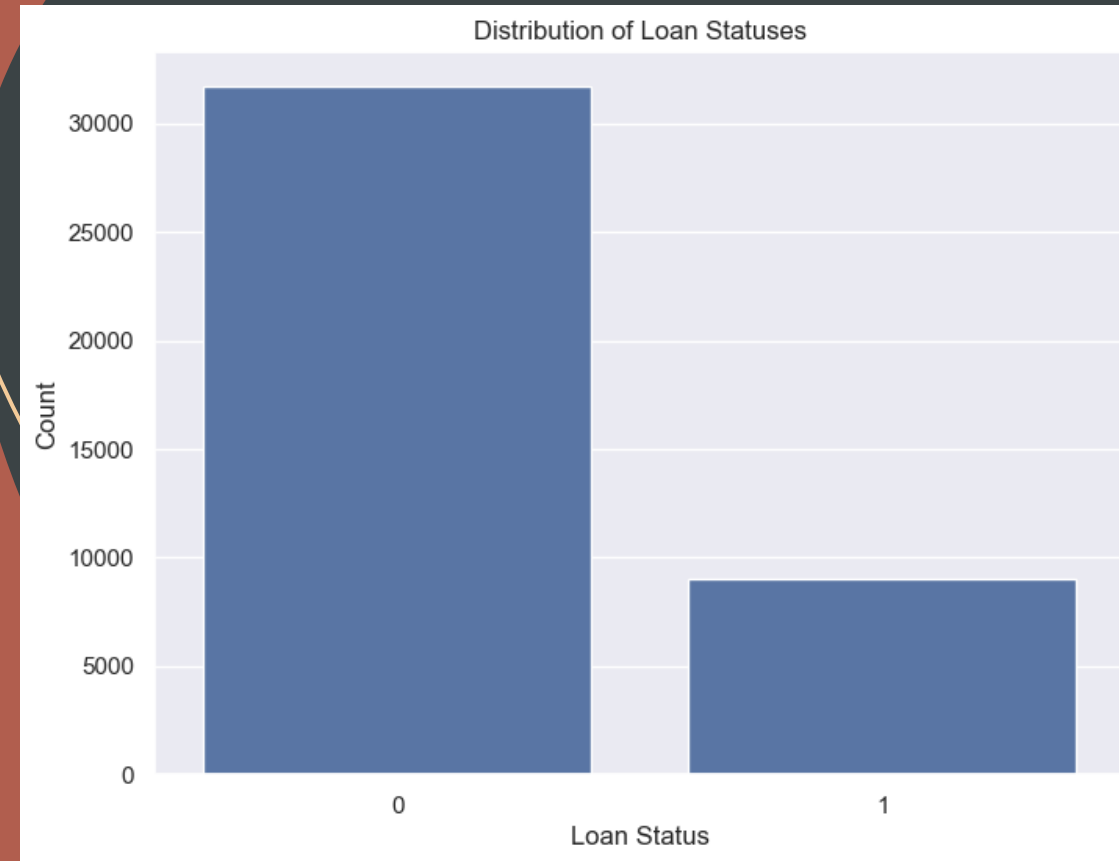
Data Cleaning

- Standardized categorical columns for better usability
- Removed outliers based on variables like loan amount and income
- This eliminated around 5,000 observations from our dataset



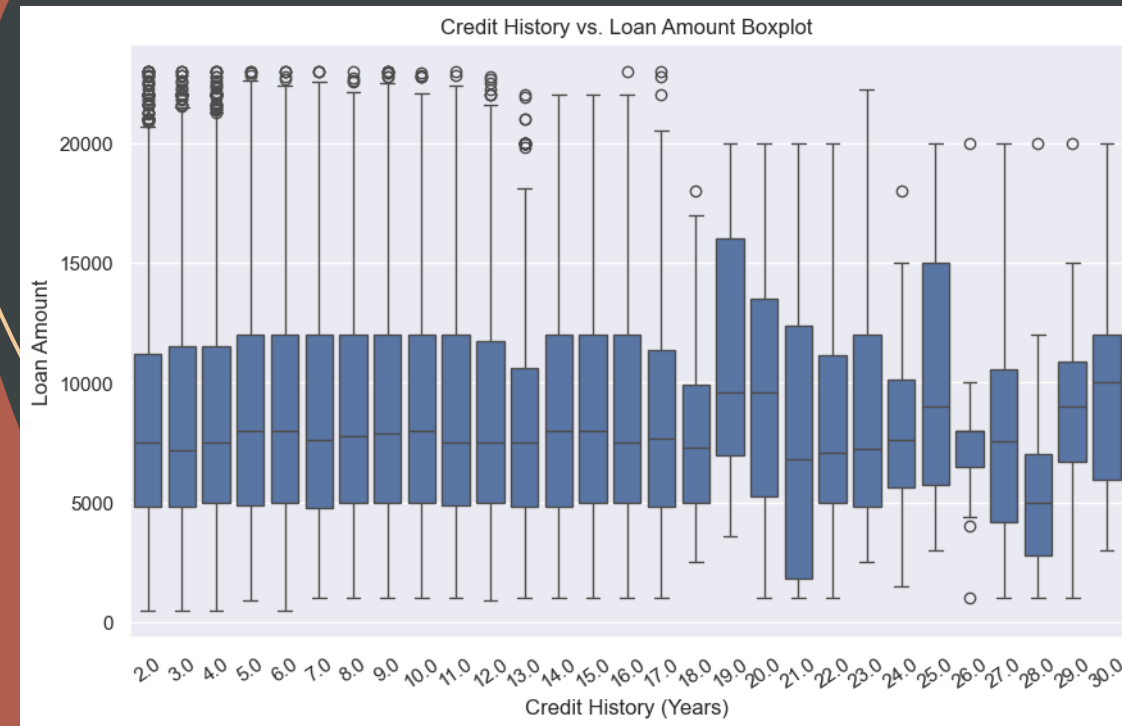
Exploratory Data Analysis

- **Distribution of loan statuses**
- Credit history vs. Loan amount
- Correlation

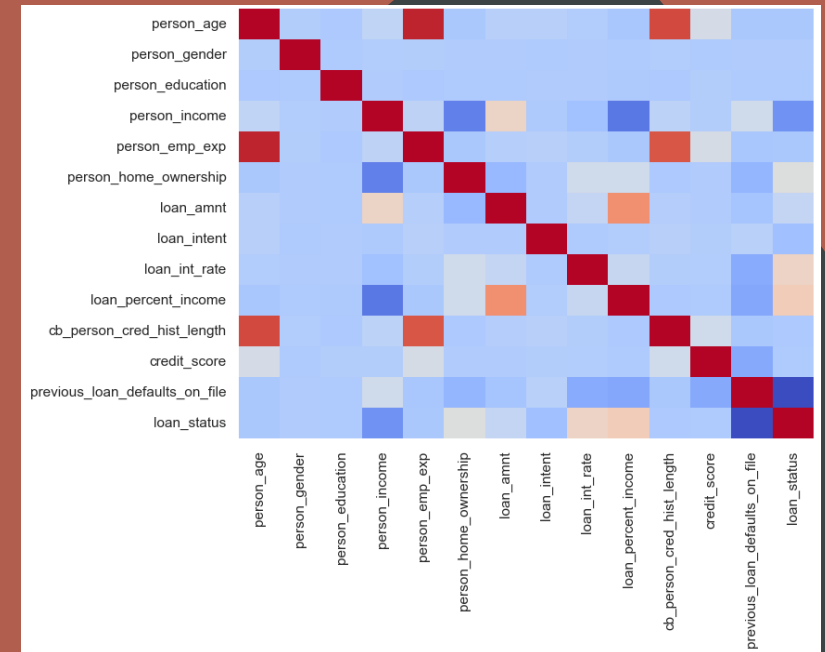


Exploratory Data Analysis cont.

- Distribution of loan statuses
- **Credit history vs. Loan amount**
- Correlation

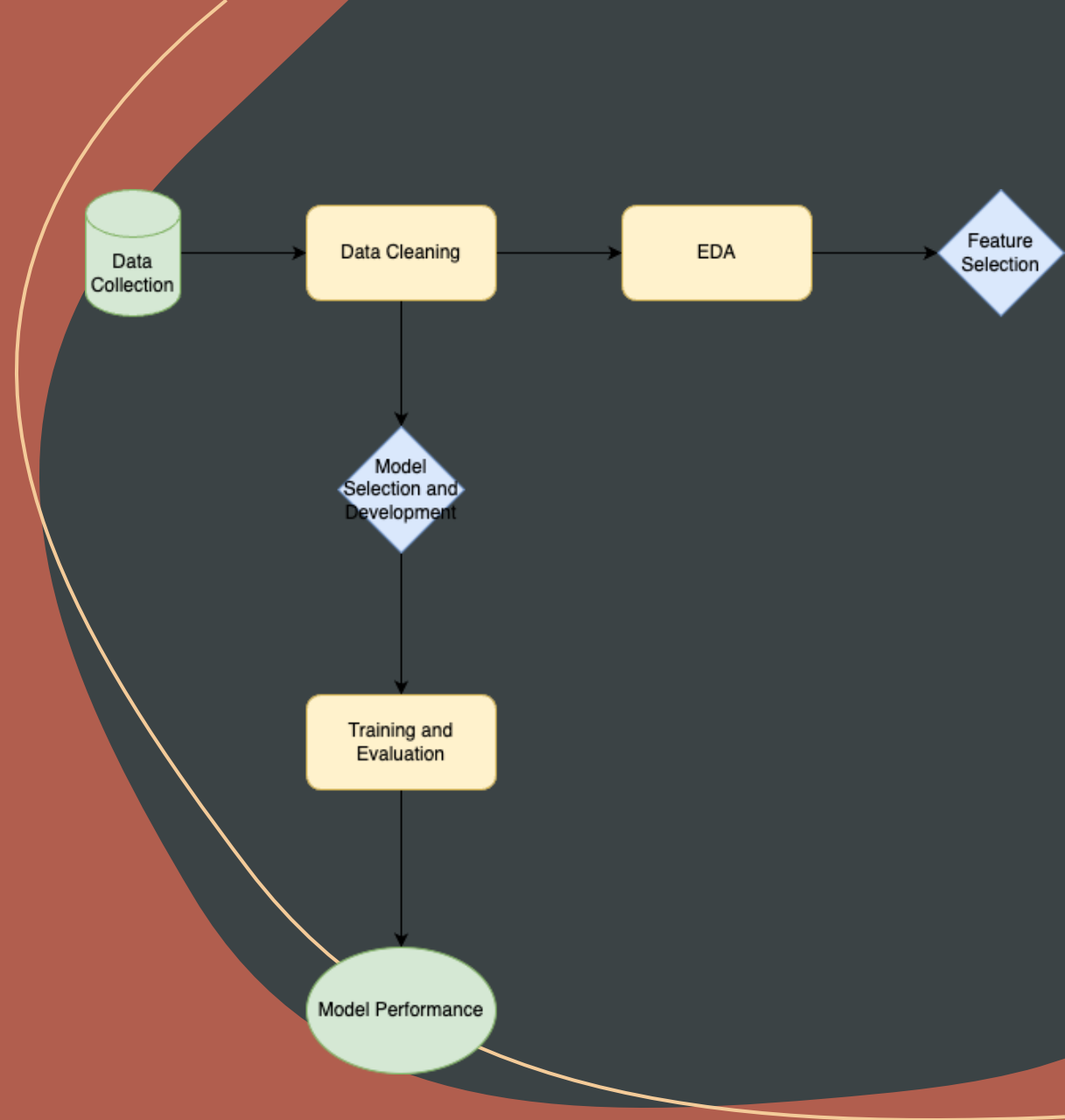


- Distribution of loan statuses
- Credit history vs. Loan amount
- **Correlation**



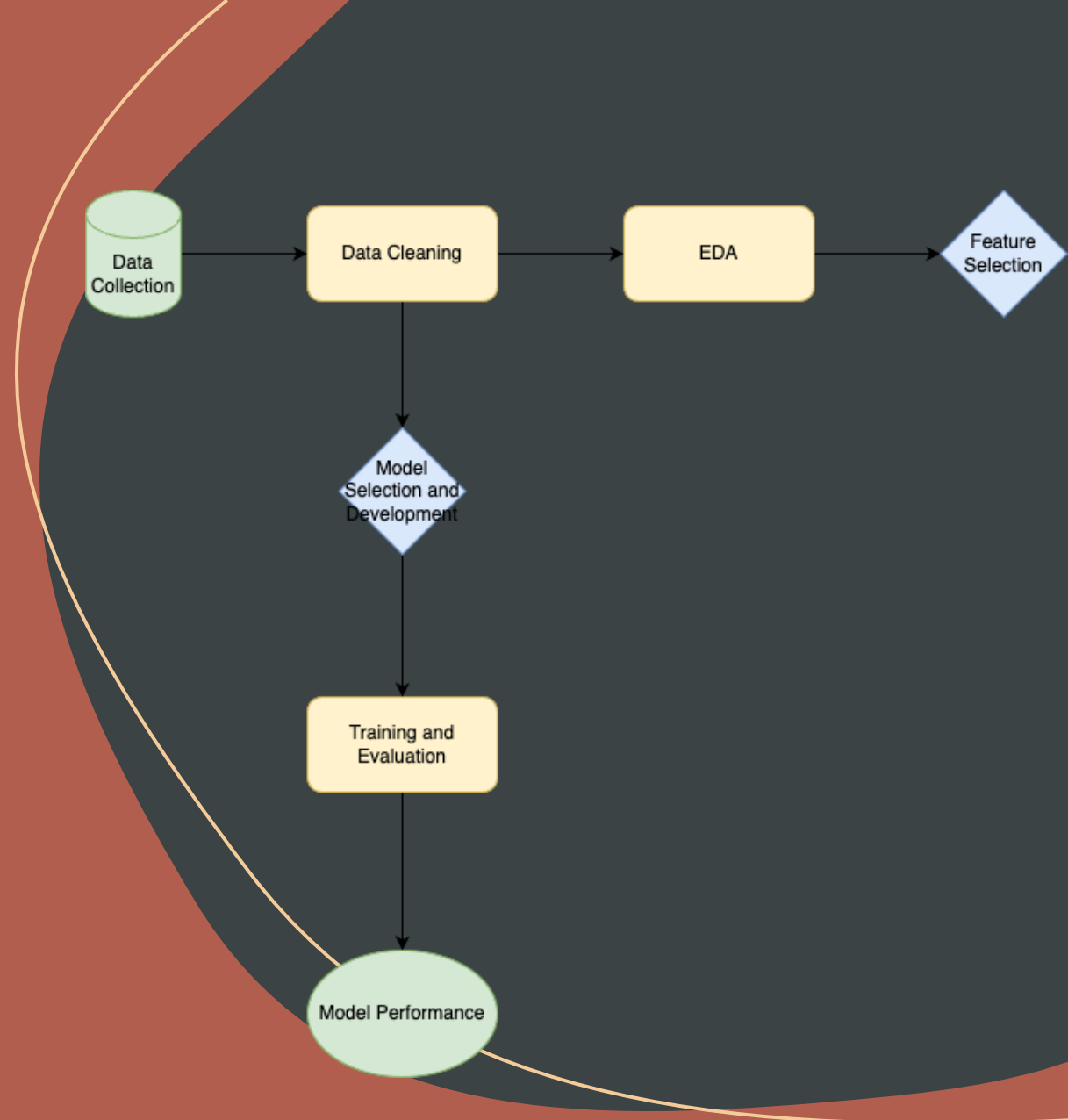
Model Development

- Decided to use Logistic Regression
- **Why?**
 - Ideal for binary classification tasks
 - Provides easy interpretability
 - Computationally efficient



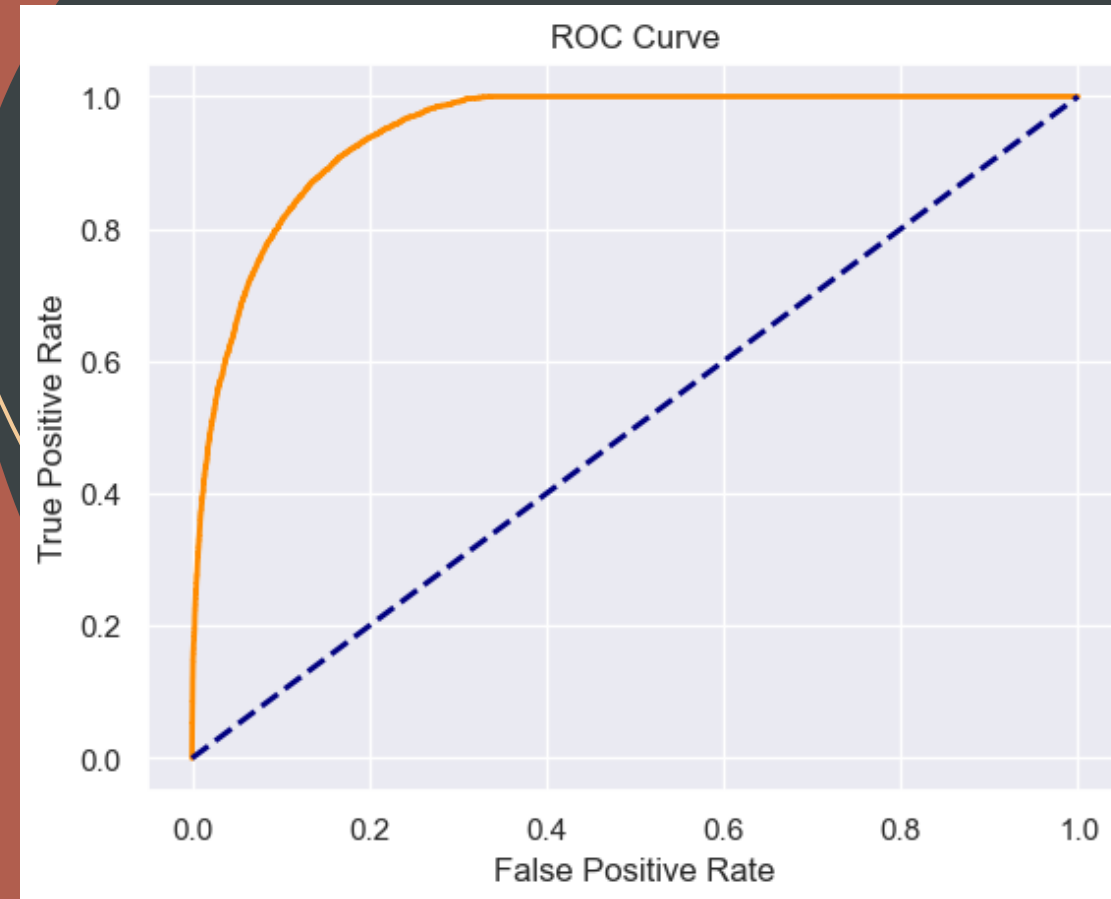
Model Development cont.

- Based on correlation analysis, we used our top 5 predictors:
 - **previous_loan_defaults_on_file**
 - **loan_percent_income**
 - **loan_int_rate**
 - **person_income**
 - **person_home_ownership**



Model Performance

- **Train Accuracy:** 89%
- **Test Accuracy:** 89%
- **Class 0 (Decline):**
 - F-score: 93%
- **Class 1 (Approve):**
 - F-score: 74%
- **Pseudo R^2 Value:** 0.331
- **AUC (Area Under Curve):** 0.946



Model Performance cont.

- **Classification Threshold: 0.5**
 - **Overall Accuracy: 88.76%**
 - **Sensitivity (Recall for Approvals): 72.62%**
 - **Specificity (Recall for Declines): 93.4%**

		Confusion Matrix	
		Predicted No	Predicted Yes
	Actual No	23665 (93.34%)	1688 (6.66%)
	Actual Yes	1972 (27.38%)	5231 (72.62%)

Summary Insights

This logistic regression model aims to reduce loan charge-offs by identifying customers likely to default

The key predictors in our model were:

Our model ended up with a 0.946 AUC and test accuracy of 88.76%

If they had any previous loans on file

The loan represented as a percentage of their income

The loan interest rate

The customer's income

The customer's homeownership type

Business Impact

Potential to improve underwriting by focusing on high-risk borrowers

Implement model to improve approval criteria and reduce net charge-offs

Future Work



Use more advanced models like Random Forest and XGBoost to improve recall on approved loans



Use real-world data for better insights and predictions



Incorporate tools like SHAP or LIME to help explain individual predictions



Thank You!

Q&A