Pythonic Minds

Ali

Hung

Linh

Sudeeptha

Table of Contents

Introduction

- Banking for the "Unbanked"
  - Not everyone has good credit

- Identifying good vs bad profiles for loans
  - Prioritize Sustainability

HOME CREDIT

"Empowering people to live the life they want now"

Overview

1. Data wrangling
   - Joining, aggregating, cleaning multiple datasets

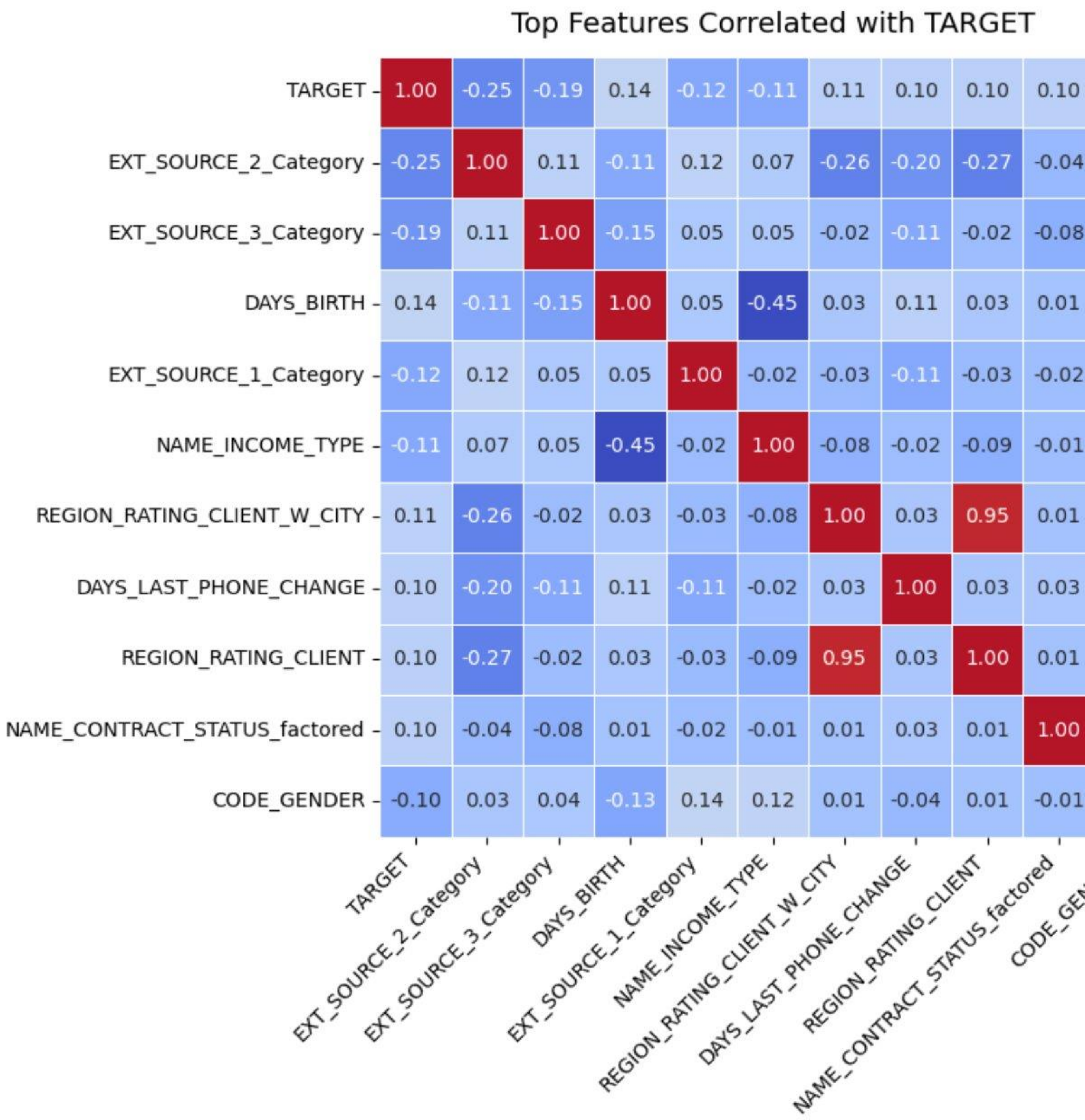
2. Modeling
   - RandomForest, KNN, Logistic Regression, CatBoost (best)

3. Key Findings
   - Confusion Matrix

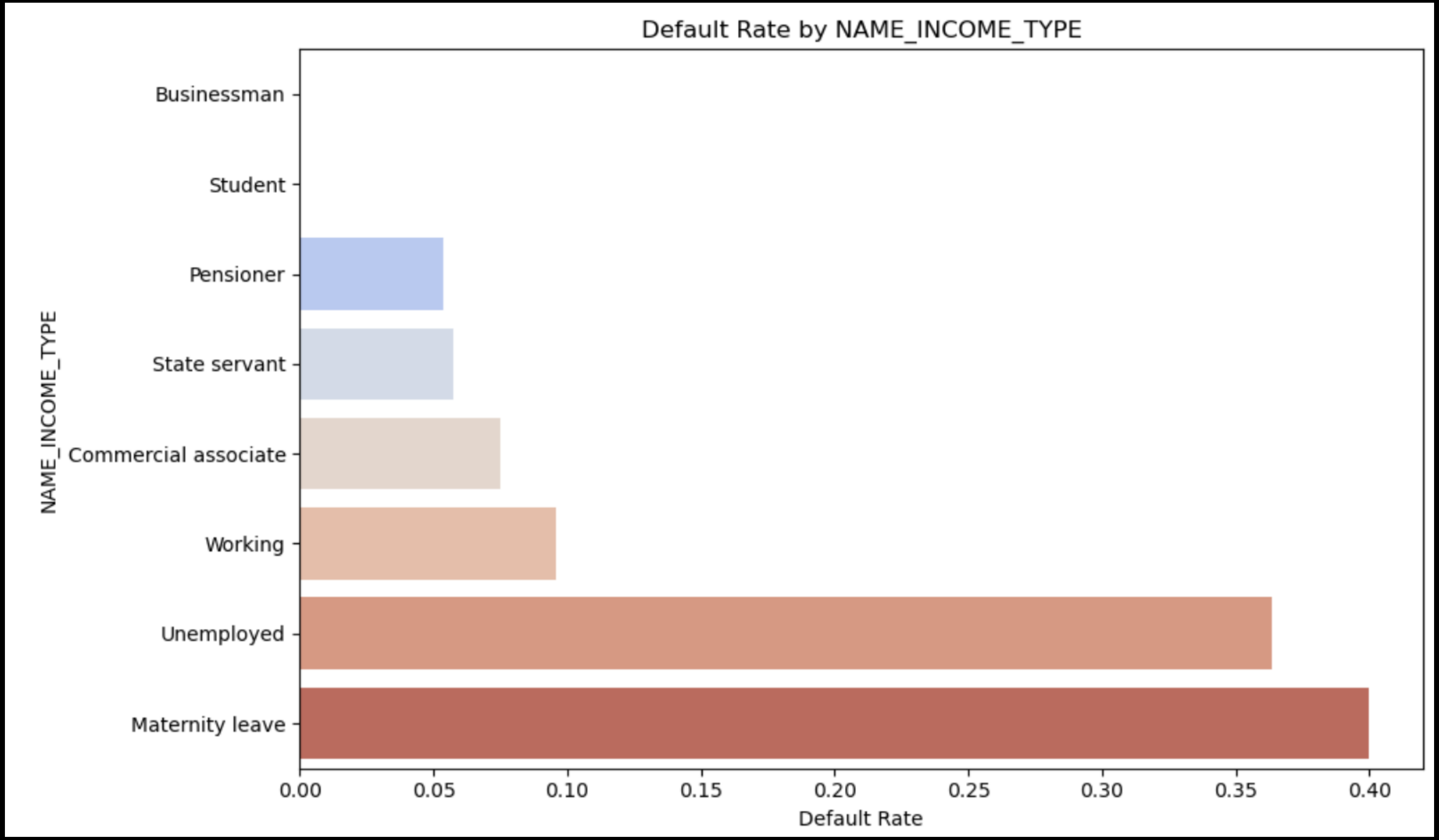
Business Insights

HOME CREDIT

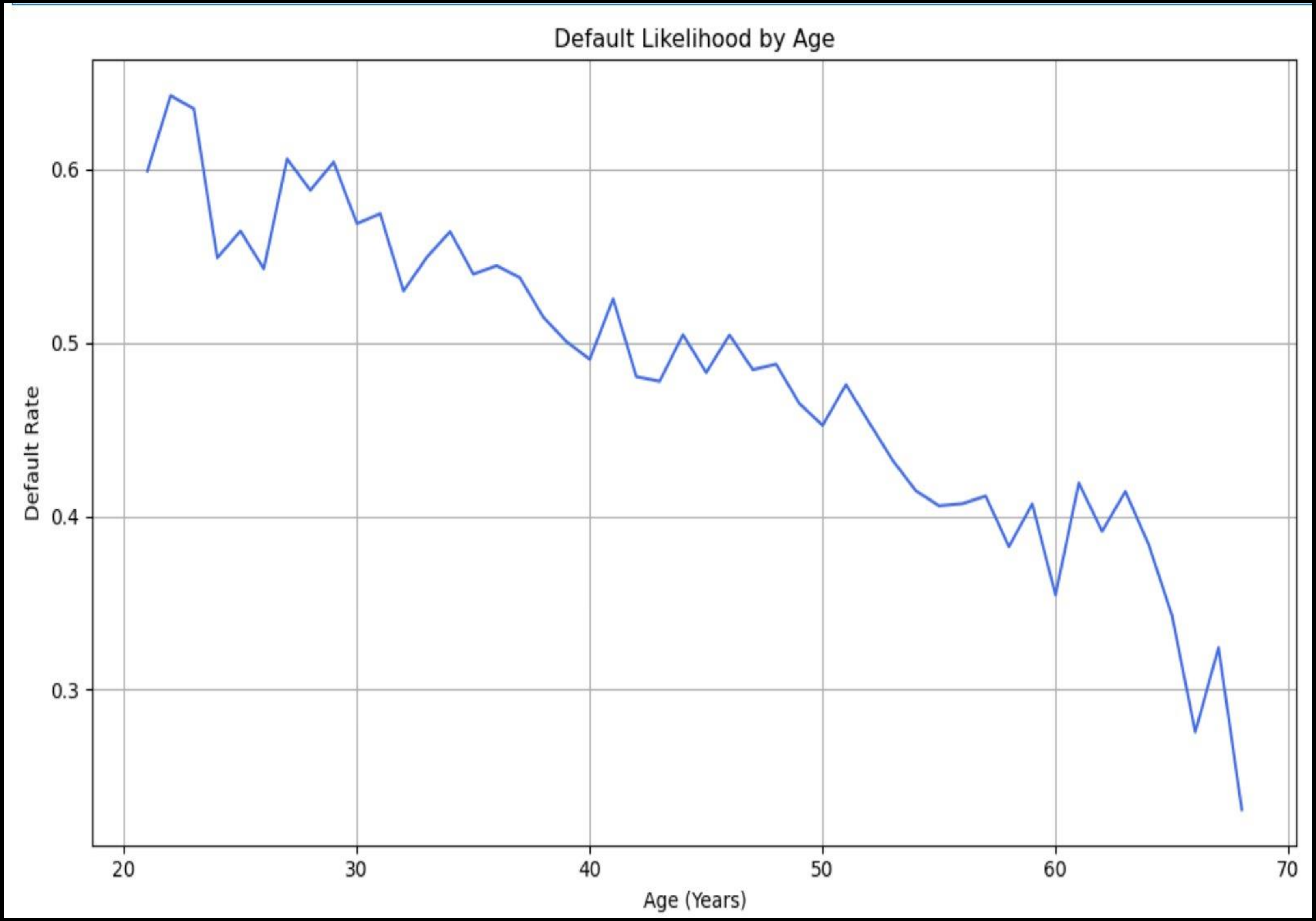
- NAME_INCOME_TYPE
- DAYS_BIRTH
- DAYS_LAST_PHONE_CHANGE
- CODE_GENDER

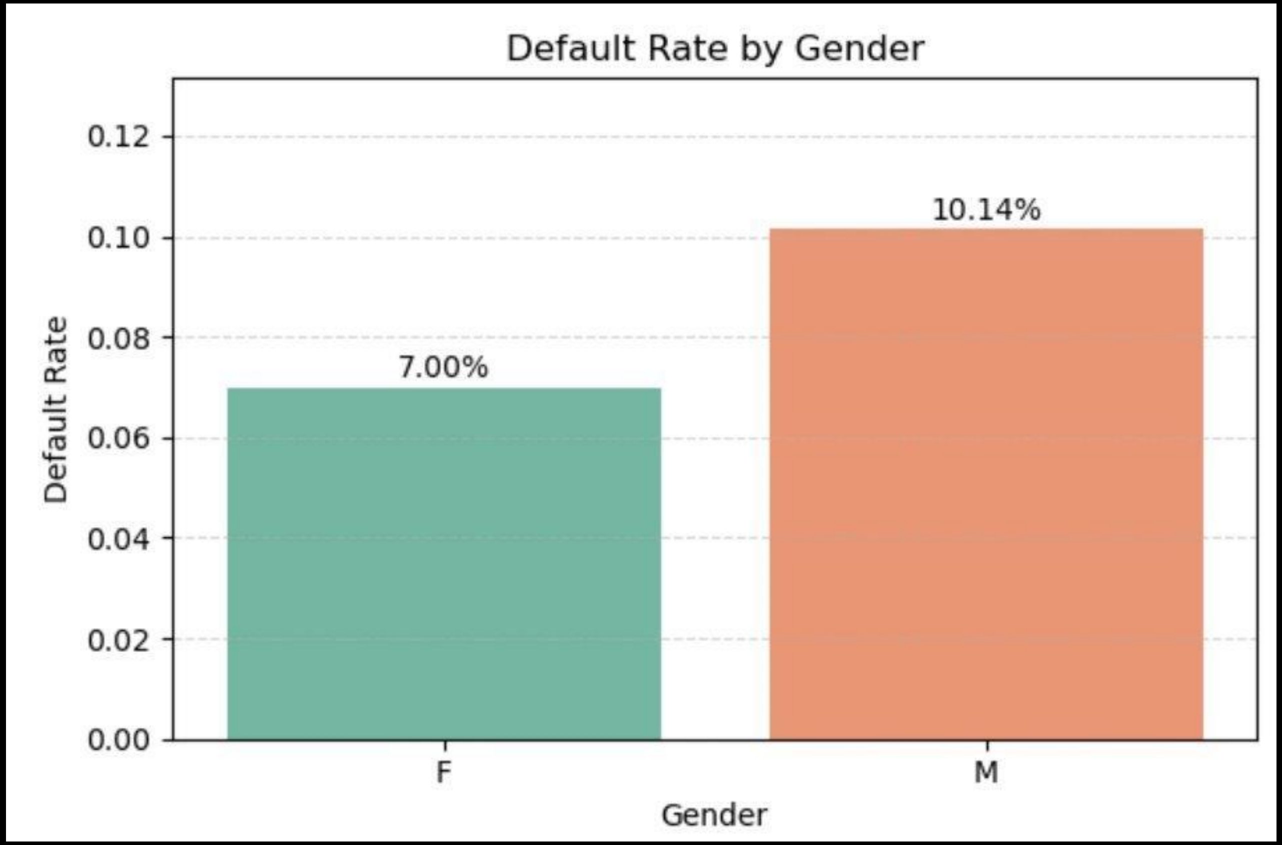


- Groups without an income tends to have has much higher risk

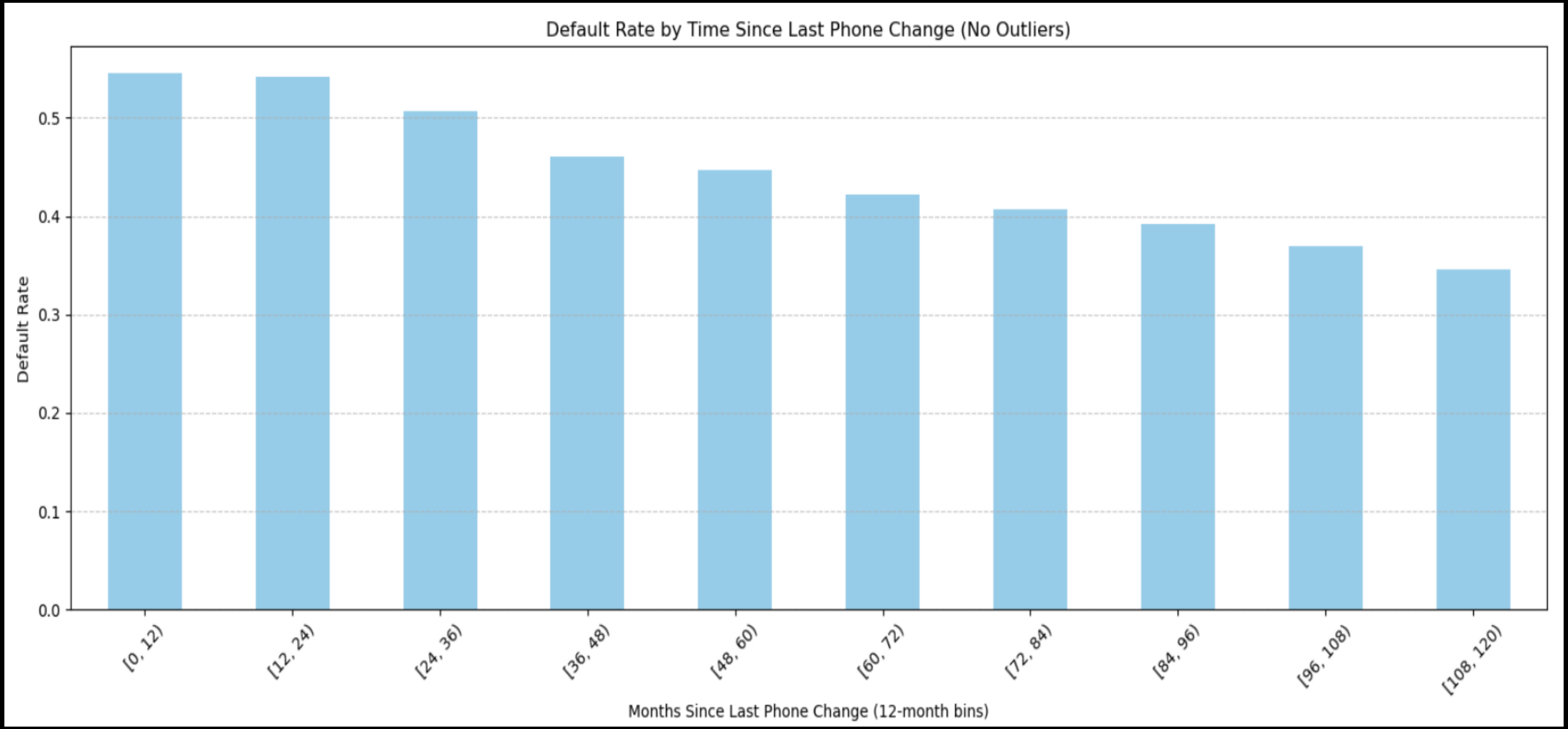


- Younger borrow groups has a higher risk





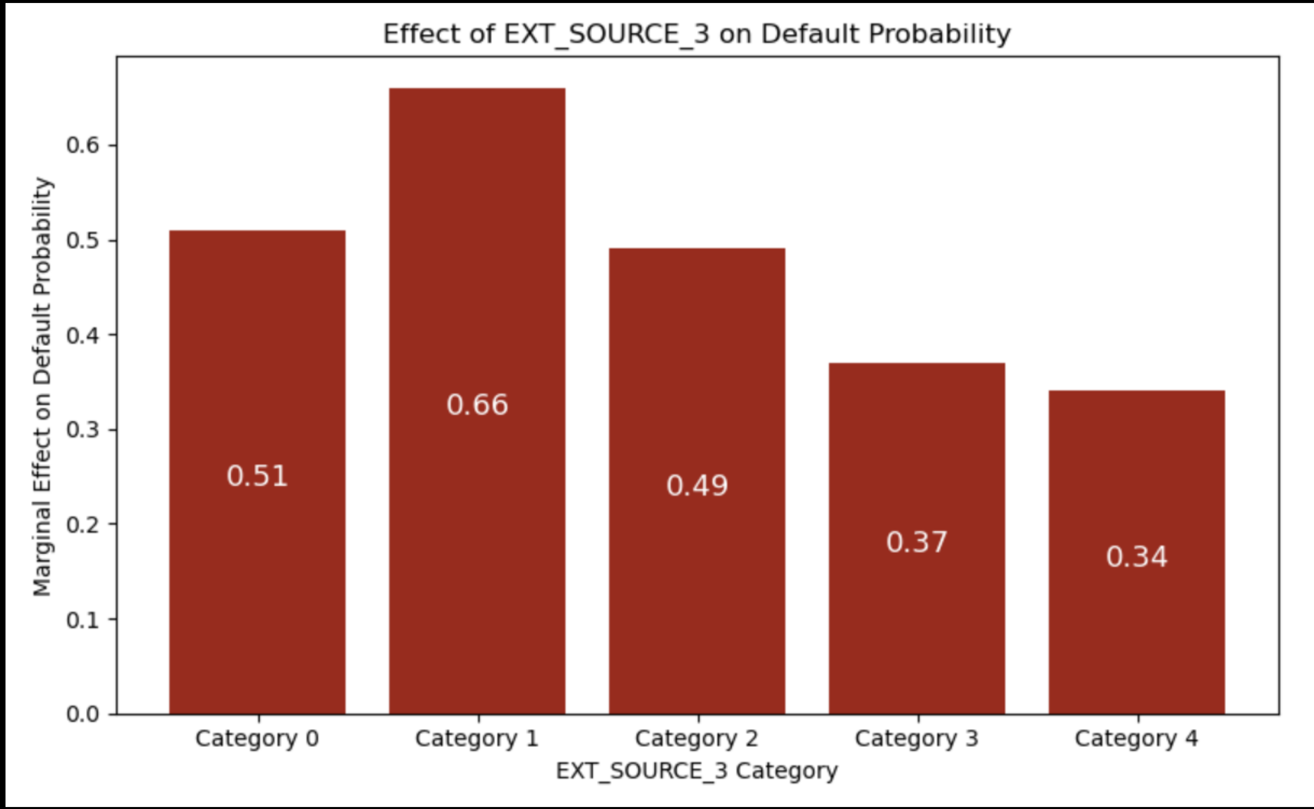
- The male group hold a higher risk than female group



- Linearly, frequency of number changing, do correlate with risk of difficulties in payment

EXT_SOURCE_3:

- The highest default risk is observed in category 1 (lowest credit score of 0 – 0.4)
- Interestingly, those with non-existent credit scores (category 0) have a lower marginal effect than category 1



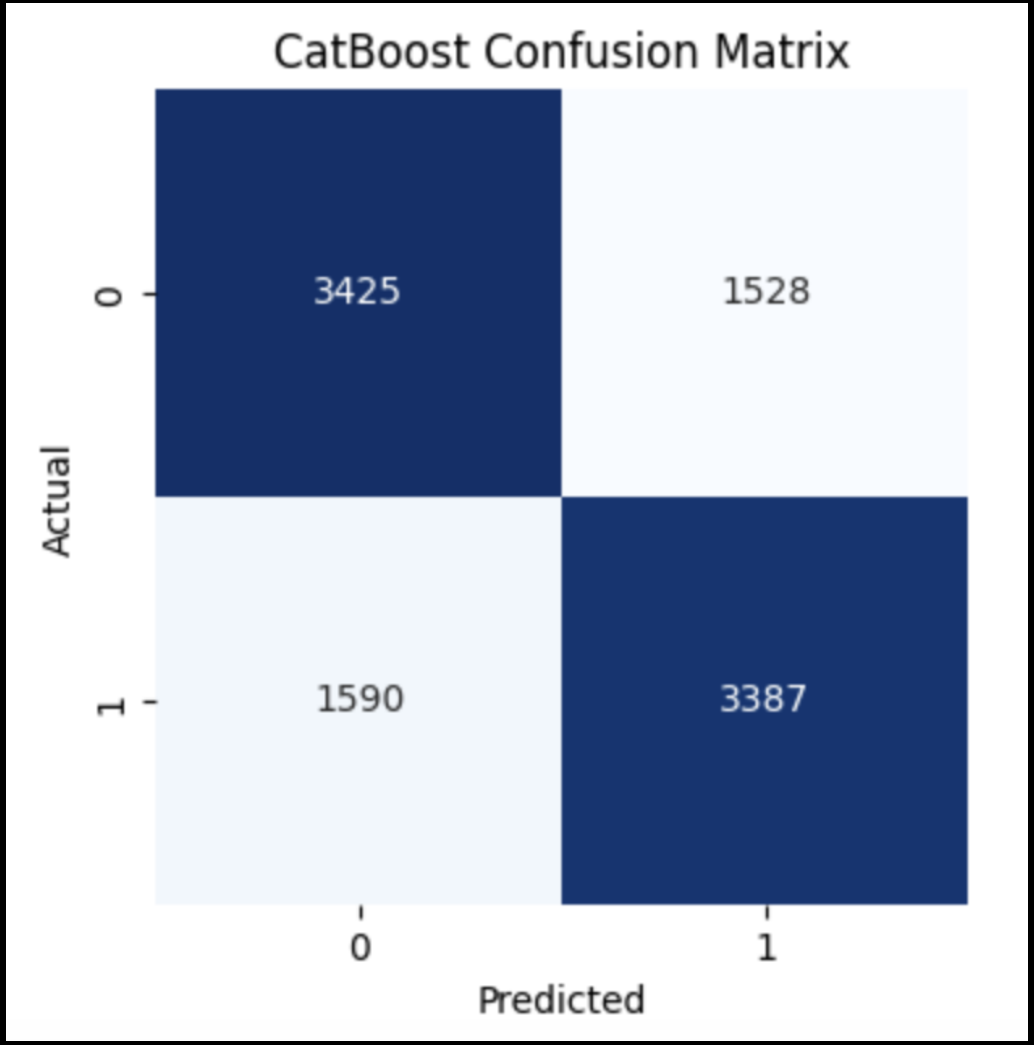
Winning Model

CatBoost

- Achieved a 70% accuracy
- Balanced predictability between both instances (default and non-defaulters)

Classification Report:

	precision	recall	f1-score	support
0	0.68	0.69	0.69	4953
1	0.69	0.68	0.68	4977
accuracy			0.69	9930
macro avg	0.69	0.69	0.69	9930
weighted avg	0.69	0.69	0.69	9930



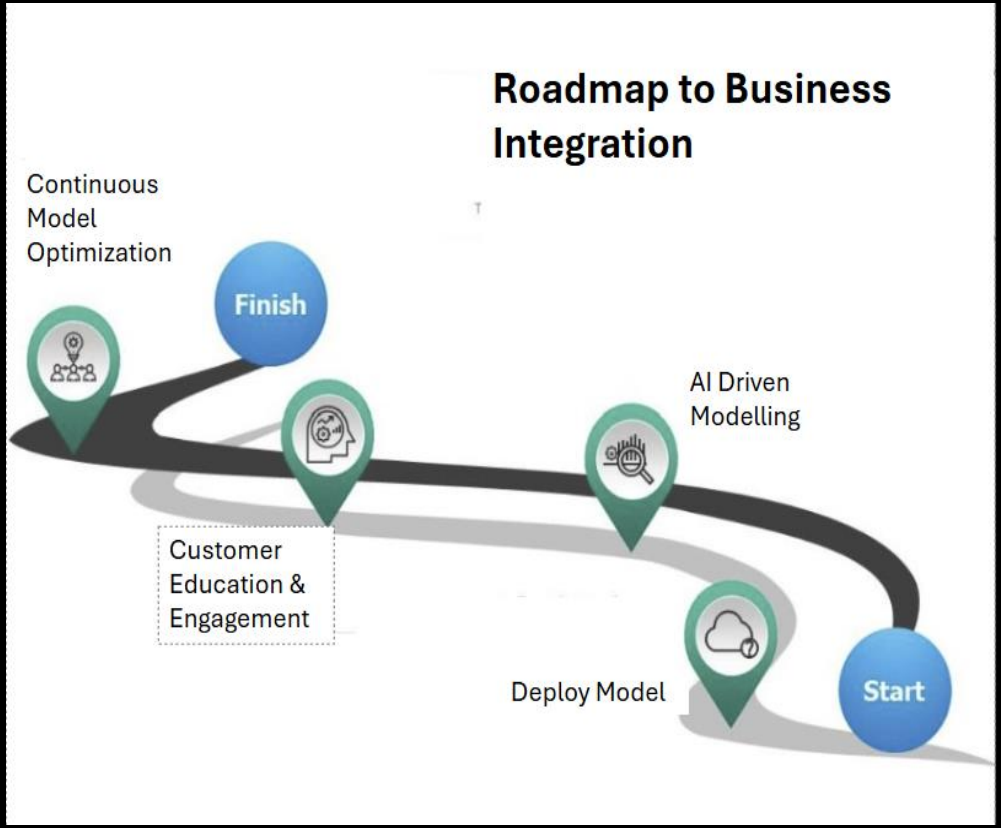
Business Solutions

- **Personalized Loan Offers:**
  - Adjusted interest rates based on predicted default probability.
- **Early Warning System:**
  - Proactively identify at-risk customers and remind them
- **Fraud Detection Enhancements:**
  - Detect suspicious loan applications to minimize losses.

Next Steps

Roadmap for Business Integration

- **Pilot Test New Model**
  - Deploy in a controlled environment to validate performance.
- **Introducing AI- Driven Models**
  - Ensure regulatory and compliance alignment.
- **Customer Education & Engagement**
  - Transparent communication about risk-based pricing.
- **Continuous Model Optimization**
  - Monitor model performance and update based on new data trends.



Thank you

Questions?