

# Pythonic Minds

Ali

Hung

Linh

Sudeeptha

# Table of Contents

1. Introduction
2. Overview
3. Business Insights
4. Model Explanation
5. Business Solutions
6. Next Steps

# 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



CatBoost

Business  
Insights

**HOME  
CREDIT**

---

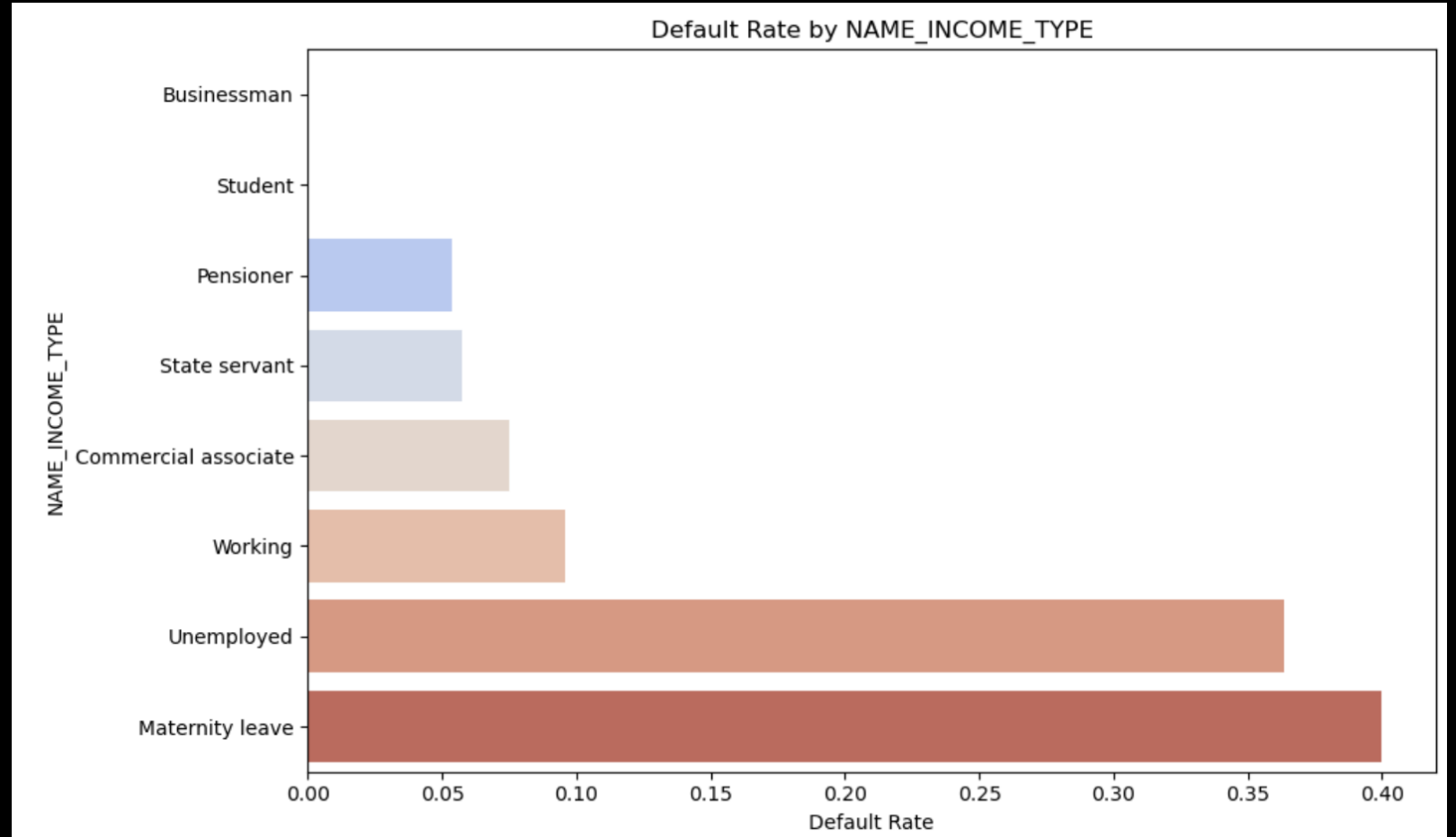


- NAME\_INCOME\_TYPE
- DAYS\_BIRTH
- DAYS\_LAST\_PHONE\_CHANGE
- CODE\_GENDER

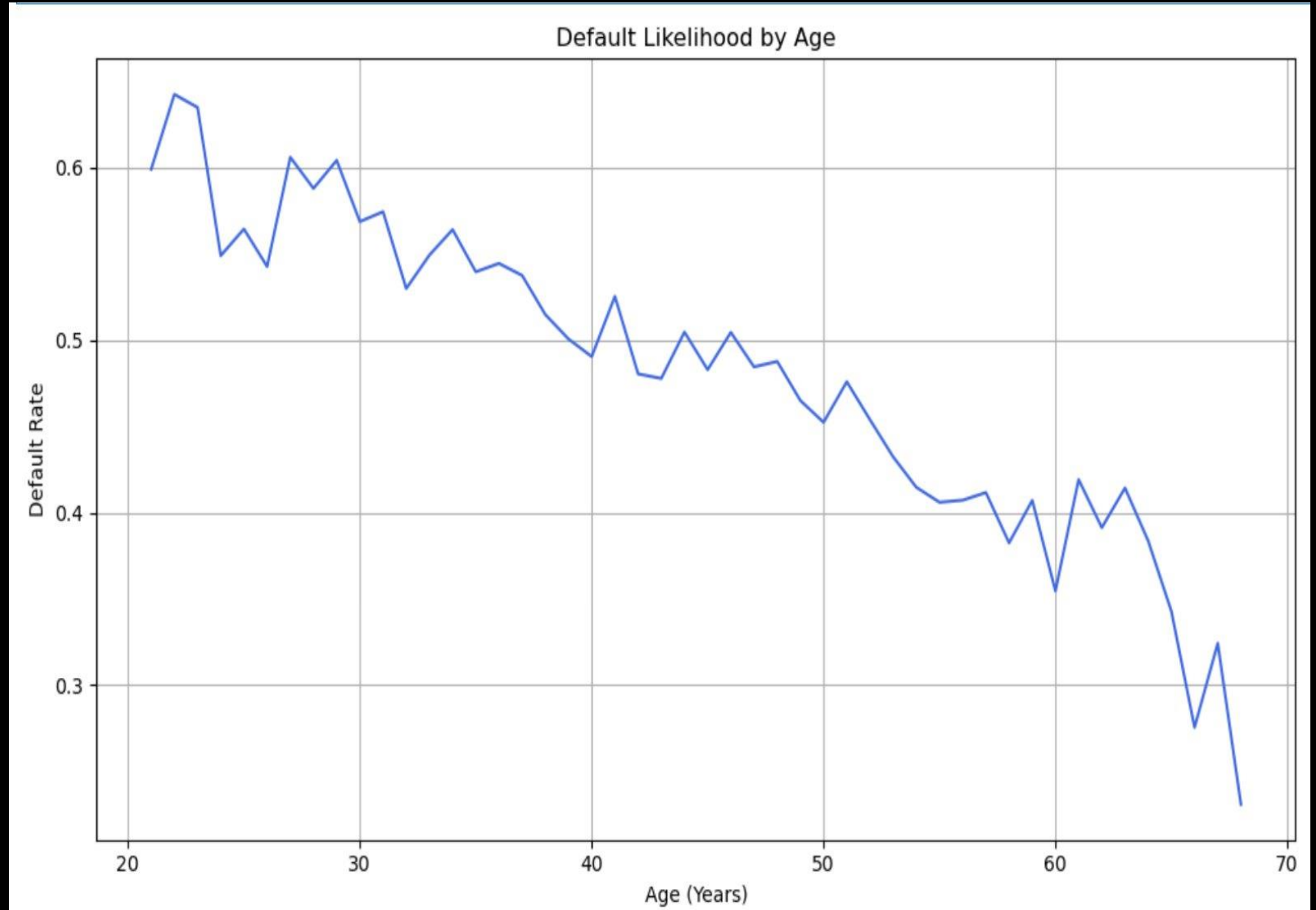
Top Features Correlated with TARGET

TARGET	1.00	-0.25	-0.19	0.14	-0.12	-0.11	0.11	0.10	0.10	0.10
EXT_SOURCE_2_Category	-0.25	1.00	0.11	-0.11	0.12	0.07	-0.26	-0.20	-0.27	-0.04
EXT_SOURCE_3_Category	-0.19	0.11	1.00	-0.15	0.05	0.05	-0.02	-0.11	-0.02	-0.08
DAYS_BIRTH	0.14	-0.11	-0.15	1.00	0.05	-0.45	0.03	0.11	0.03	0.01
EXT_SOURCE_1_Category	-0.12	0.12	0.05	0.05	1.00	-0.02	-0.03	-0.11	-0.03	-0.02
NAME_INCOME_TYPE	-0.11	0.07	0.05	-0.45	-0.02	1.00	-0.08	-0.02	-0.09	-0.01
REGION_RATING_CLIENT_W_CITY	0.11	-0.26	-0.02	0.03	-0.03	-0.08	1.00	0.03	0.95	0.01
DAYS_LAST_PHONE_CHANGE	0.10	-0.20	-0.11	0.11	-0.11	-0.02	0.03	1.00	0.03	0.03
REGION_RATING_CLIENT	0.10	-0.27	-0.02	0.03	-0.03	-0.09	0.95	0.03	1.00	0.01
NAME_CONTRACT_STATUS_factored	0.10	-0.04	-0.08	0.01	-0.02	-0.01	0.01	0.03	0.01	1.00
CODE_GENDER	-0.10	0.03	0.04	-0.13	0.14	0.12	0.01	-0.04	0.01	-0.01

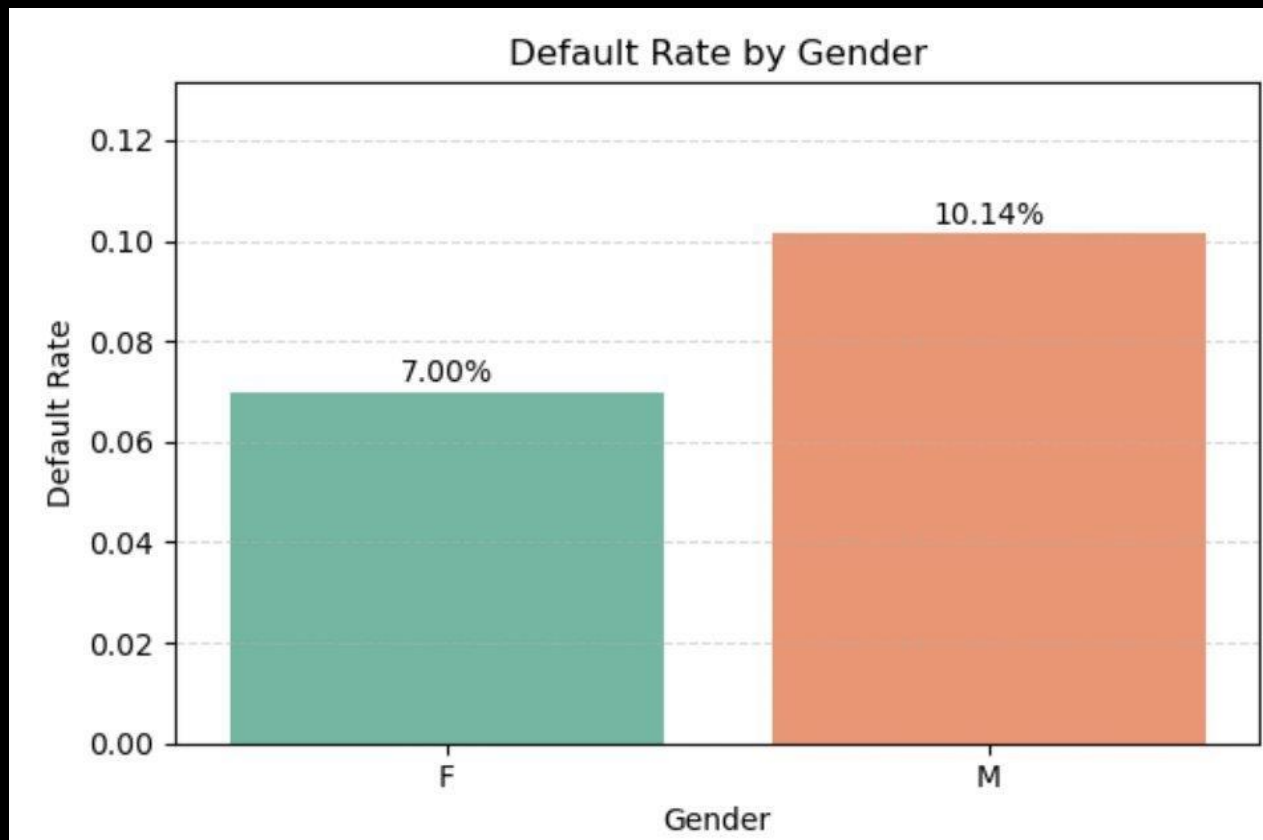
- Groups without an income tends to have has much higher risk



- Younger borrow groups has a higher risk

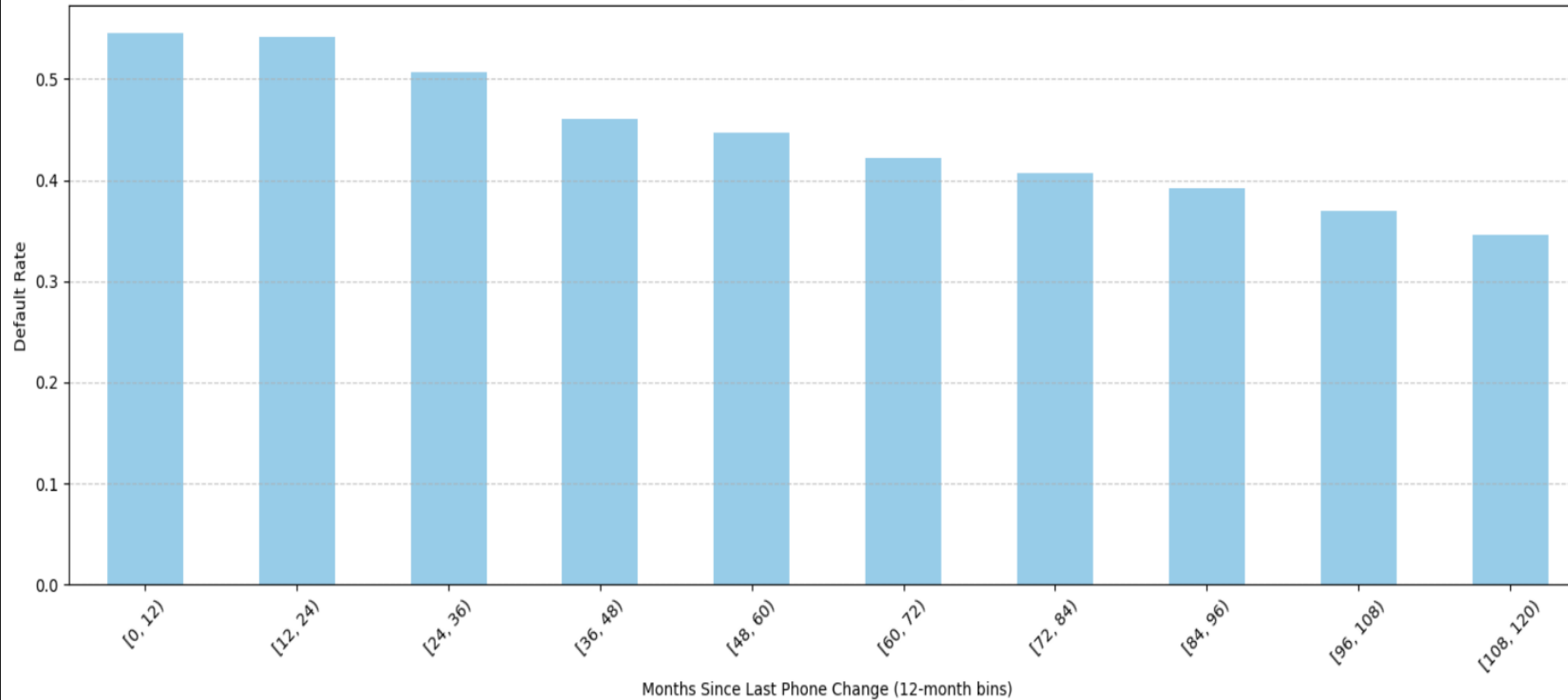






- The male group hold a higher risk than female group

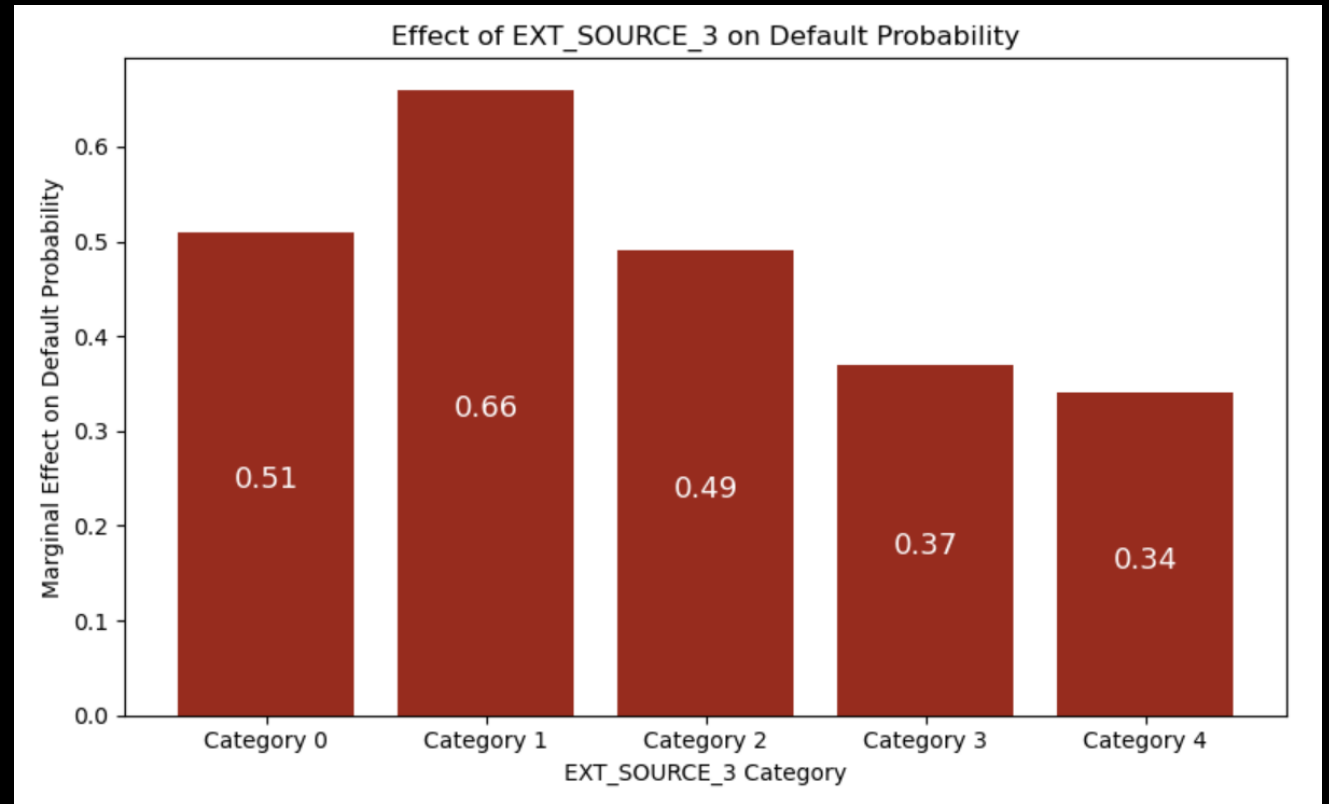
Default Rate by Time Since Last Phone Change (No Outliers)



- 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

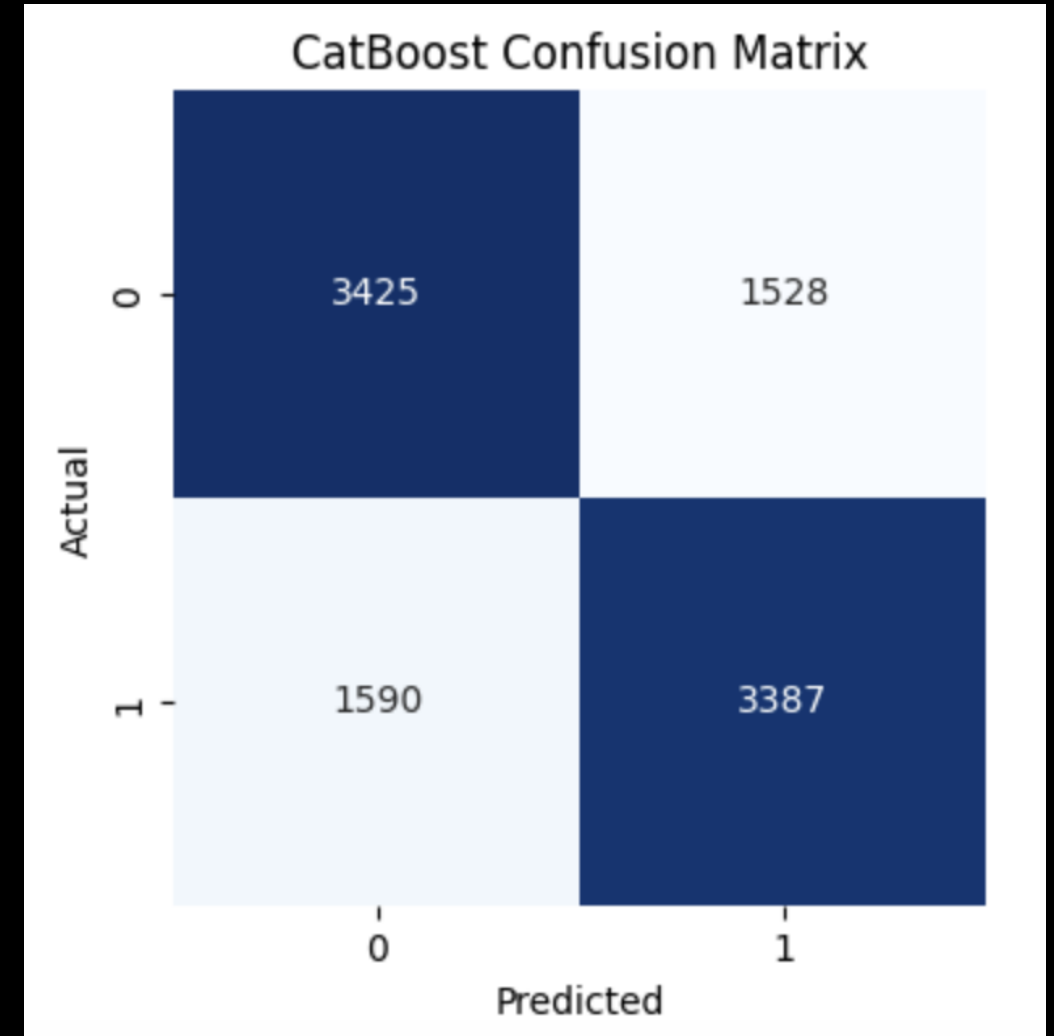


# Winning Model

## 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

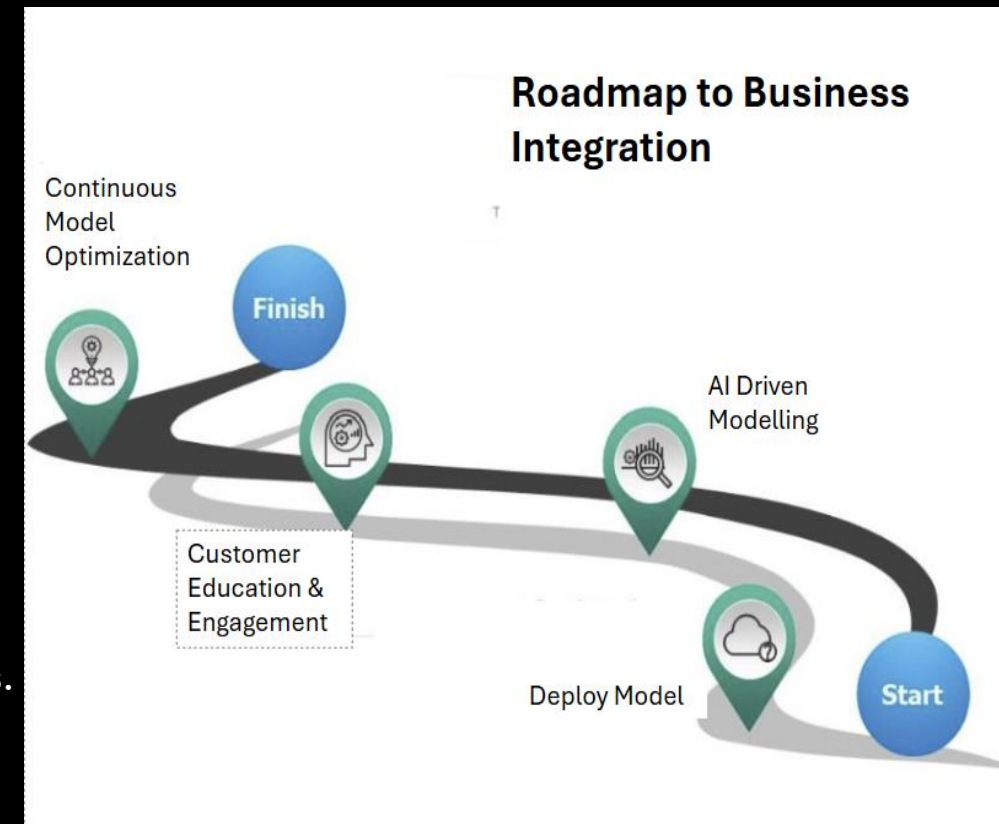
- **Personalized Loan Offers:**
  - Adjusted interest rates based on predicted default probability.
- **Early Warning System:**
  - Proactively identify at-risk customers and remind them
- **Fraud Detection Enhancements:**
  - Detect suspicious loan applications to minimize losses.



# 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?

