HOME CREDIT
LOAN APROVAL ANALYSIS

Presented By :

Aldi Vibriani

Aldi Vibriani

Project link github_Repository

AGENDA

25%

BACKGROUND PROJECT

Background of the project related to loan approval in Home Credit.

50%

EXPLORATORY DATA ANALYSIS

Understanding what happened and how to solve the problem.

75%

TECHNICAL ASPECT

Data preprocessing, feature extraction, modelling, tuning, and feature importance.

100%

BUSINESS RECOMMENDATION

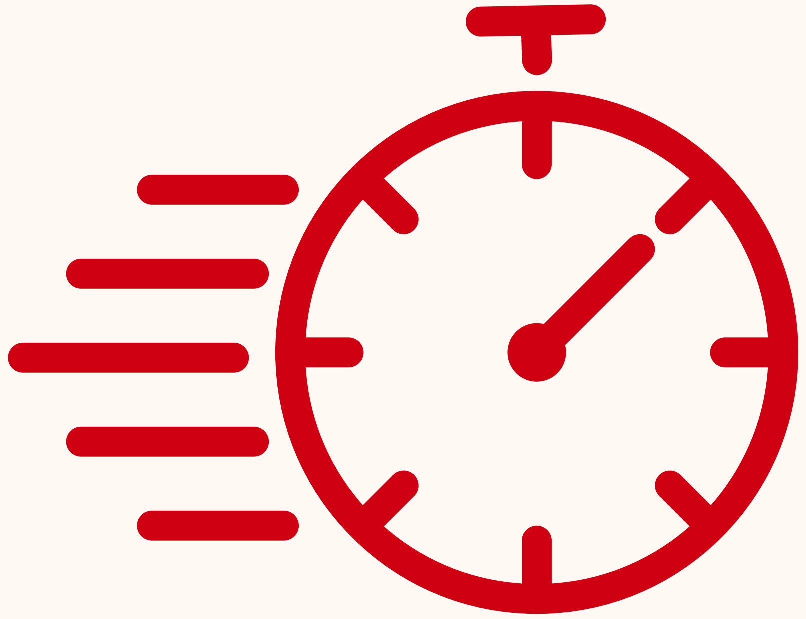
Business recommendation to predict rejected and approved loan.

BACKGROUND PROJECT

Home Credit is a consumer finance company that aims to provide individuals with access to necessary products and service. Focusing on convenience and speed, Home Credit offers credit solutions that enable customers to make purchases without having to pay the full amount upfront.

Differences between Home Credit and banks:

- **Quick Loan Approval:** The process is fast, saving both time and effort. Loans are often approved with little delay.
- **Flexible Repayment Options:** Borrowers have several repayment methods to choose from, simplifying loan management.
- **No Collateral Required:** Home Credit offers unsecured loans, allowing individuals to access funds without needing collateral.



BACKGROUND PROJECT

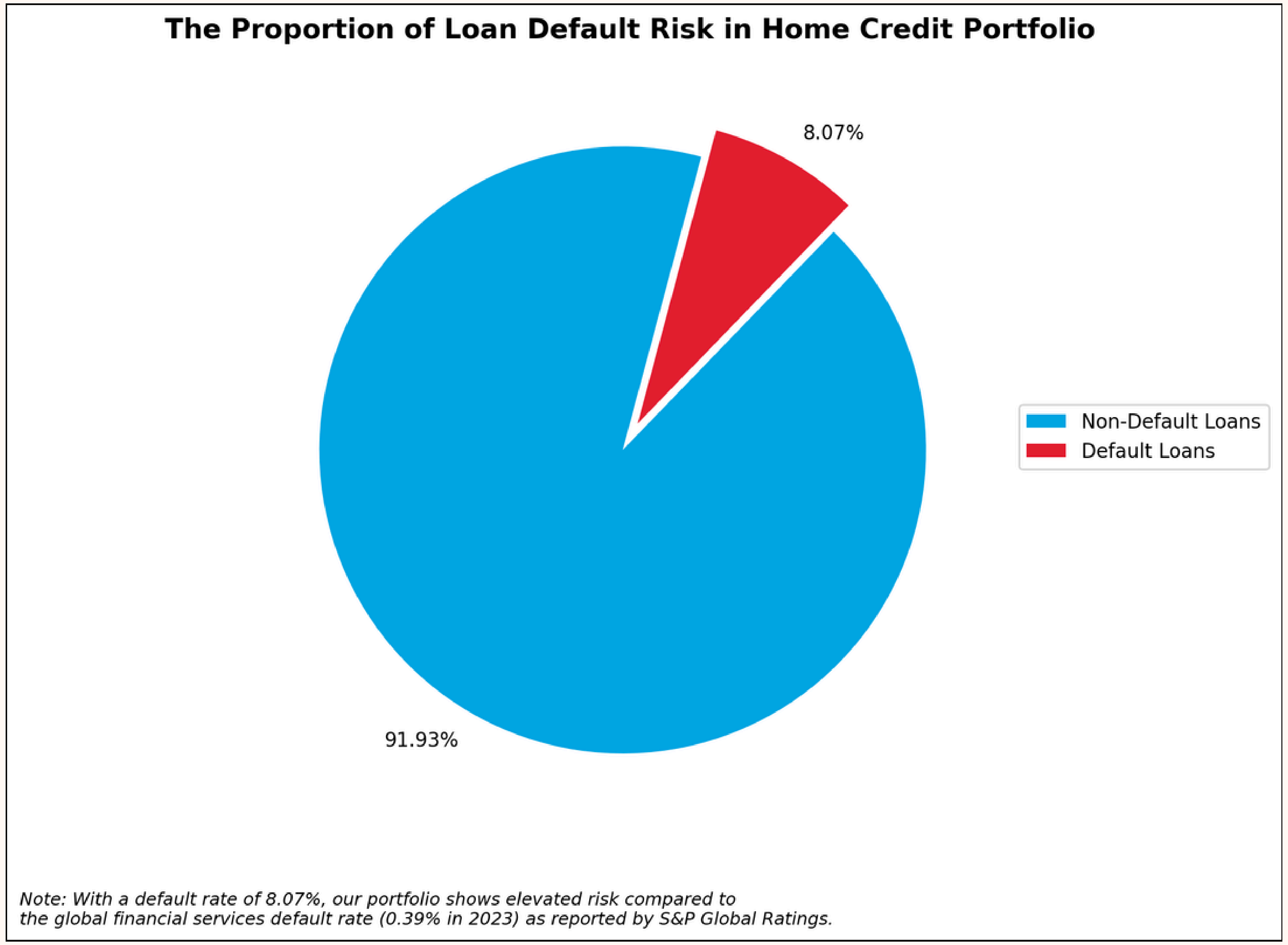


Image 1.1 - Default vs Non-Default

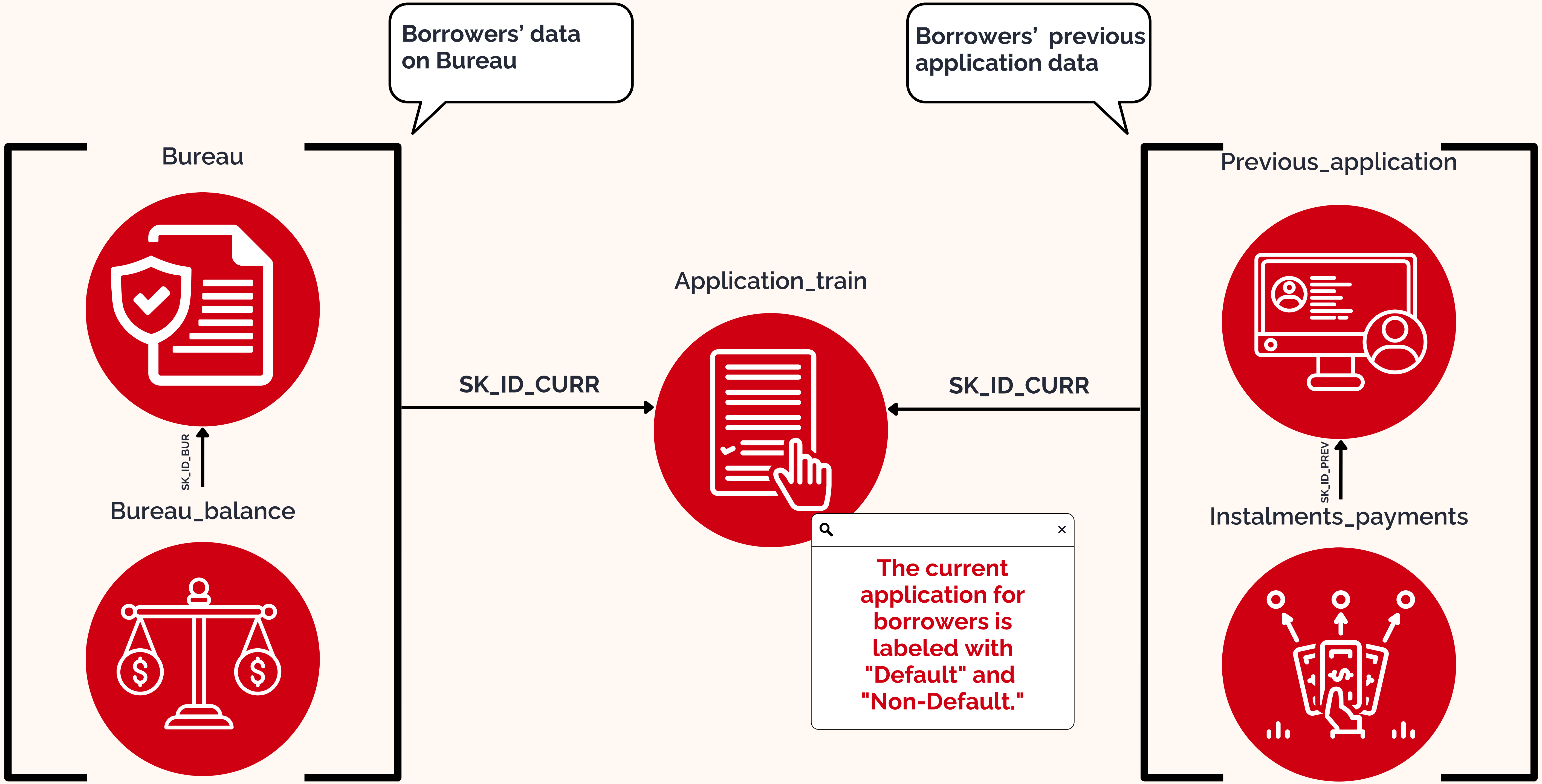
The default rate of 8.07% is substantially higher than the default rates observed in any of the sectors reported by S&P Global Ratings, it is clear that this rate is considered very high risk. Therefore, it would be prudent for Home Credit to reduce its default rate significantly to align more closely with industry averages and reduce the associated risk.

Using machine learning to reduce default rates offers several significant advantages over traditional methods, making it a more effective and efficient approach for several reasons:

1. **Handling Large Datasets:** Machine learning can process vast amounts of data, including diverse variables such as demographics, employment history, credit behavior, and other relevant factors. (Lin, J. 2024)
2. **Real-Time Predictions and Updates:** Machine learning models can be trained and updated in real-time, allowing for continuous improvement in predictive accuracy as new data becomes available. (fundmore.ai)
3. **Efficiency and Scalability:** Machine learning automates the process of credit risk assessment, reducing the need for manual intervention and the time required to analyze loan applications. (fundmore.ai)

EXPLORATORY DATA ANALYSIS

We have eight datasets available for modeling, but due to the limitations of our computer's performance, we will only utilize five of them. These five datasets include: Application_train (our primary dataset), bureau, bureau_balance, previous_application, and instalments_payments.



BUSINESS INSIGHT

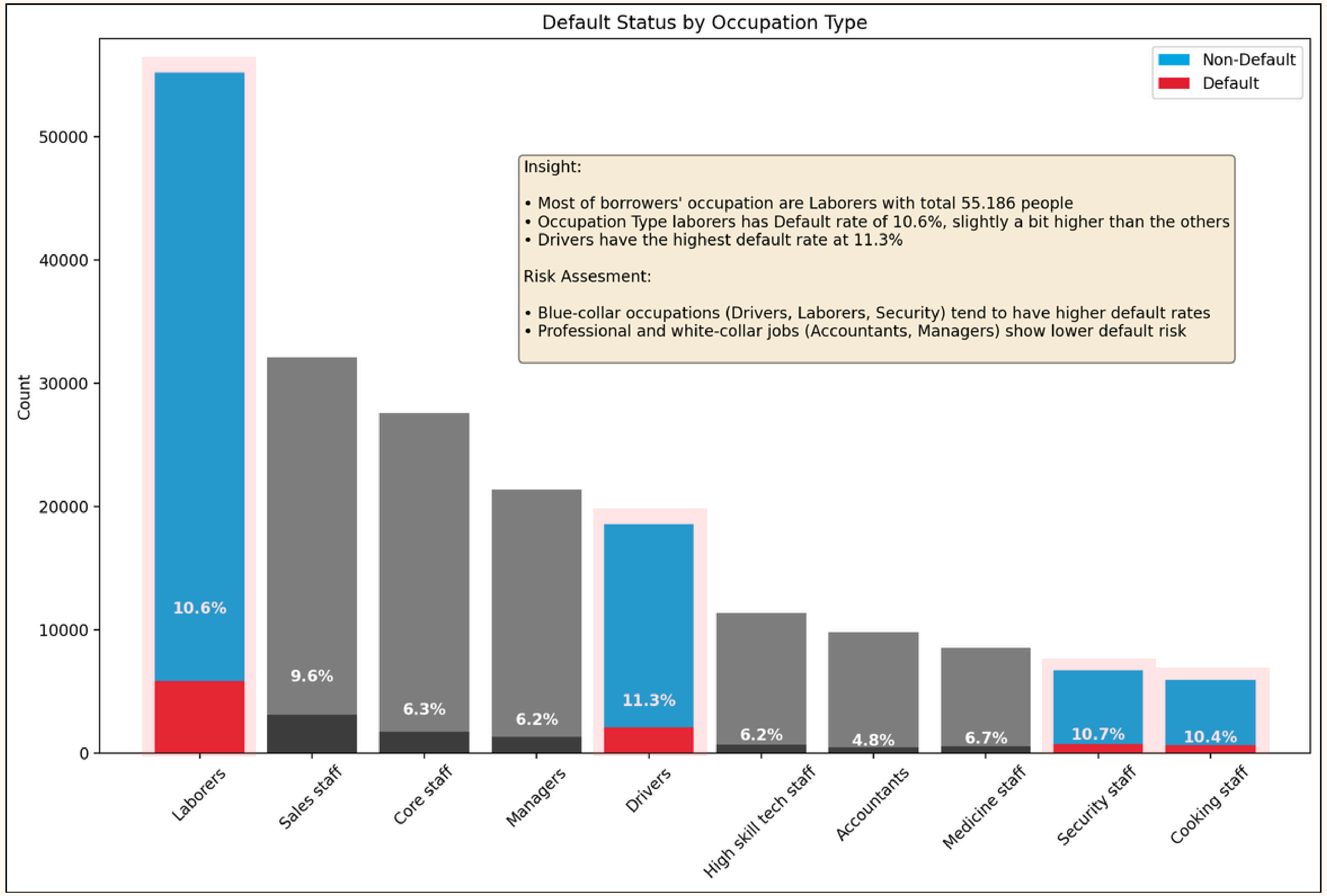


Image 1.3 - Default Status by Occupation Type

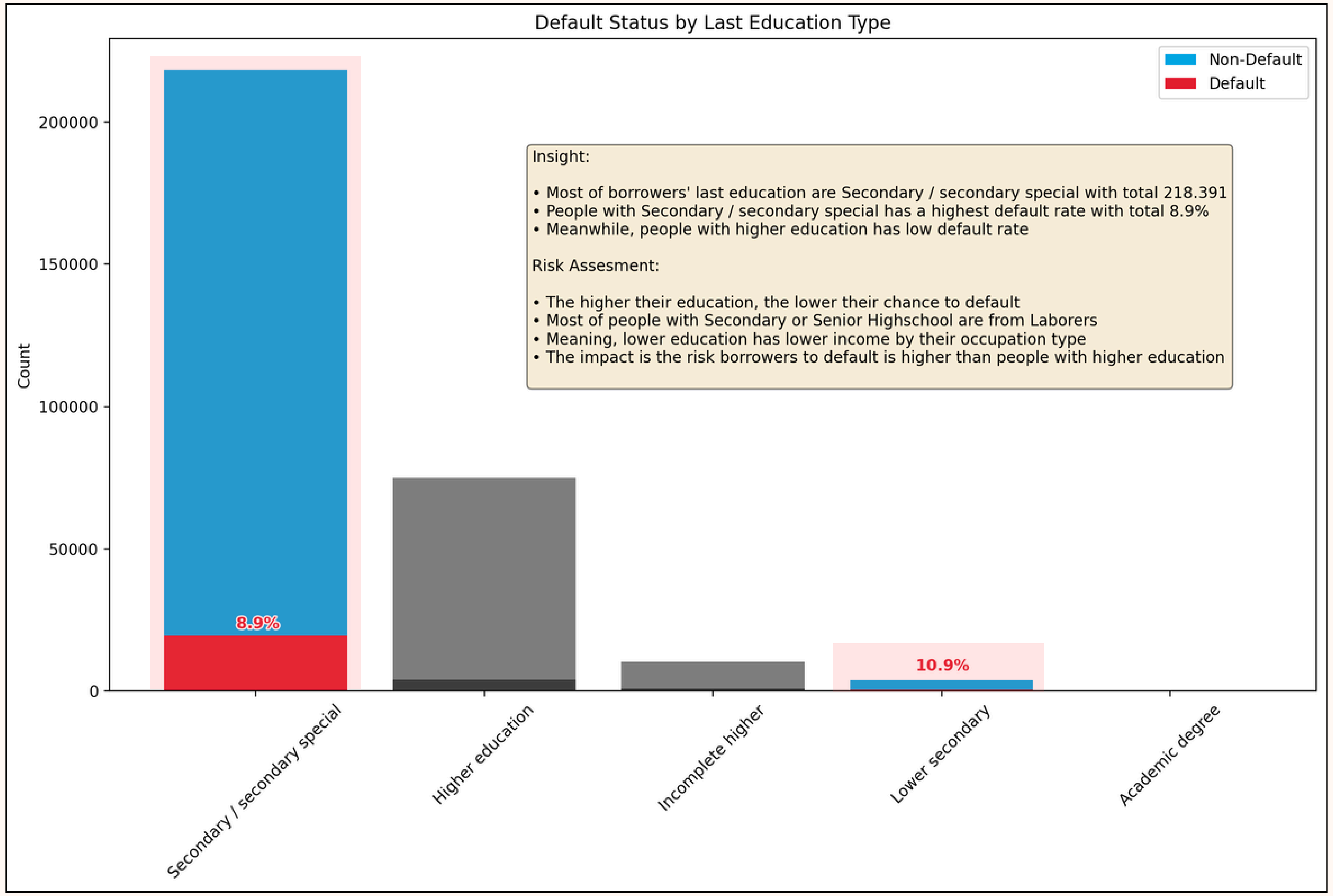


Image 1.4 - Default Status by Last Education

The majority of our borrowers, totaling 218,391 individuals, have a secondary or secondary special education background (equivalent to senior high school). Among these, laborers represent a significant portion, with 55,186 borrowers. Occupations such as laborers, drivers, security staff, and cooking staff predominantly have a secondary education level and exhibit a higher default rate, exceeding 10%.

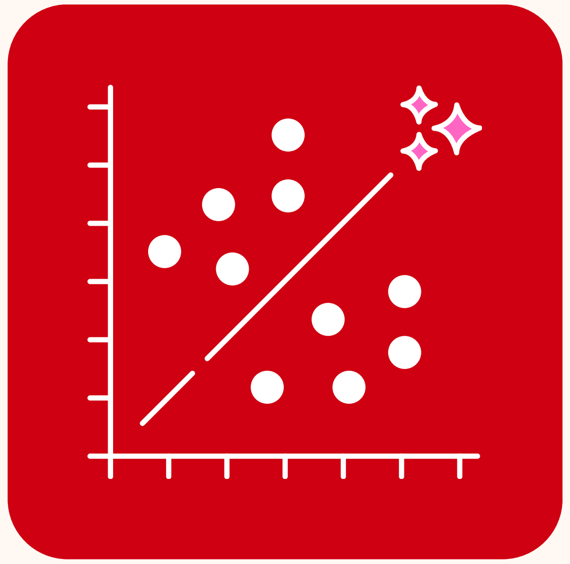
This indicates that borrowers with lower educational attainment present a higher risk of default. Conversely, those with higher education, such as accountants, high-skilled tech staff, or managers—who typically hold higher education or academic degrees—demonstrate a lower default rate. This trend is illustrated in Image 1.3.

DATA CLEANING

HANDLING MISSING VALUE

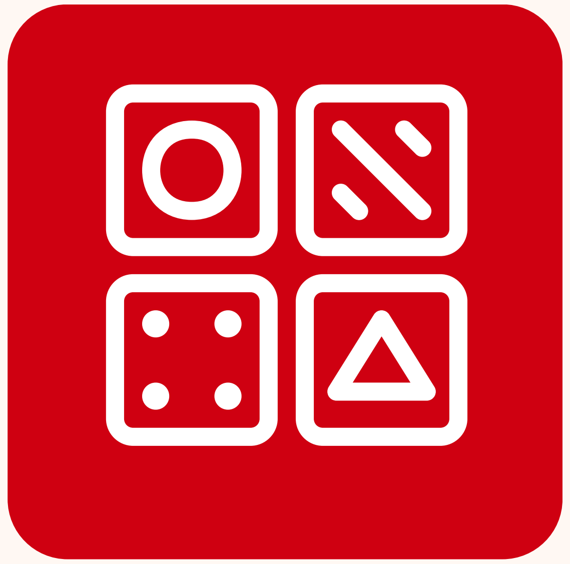
There are many columns with missing values. We handling it with data imputation and drop columns / rows.

HANDLING OUTLIERS



Some columns contain outliers, with some being anomalies and others not. For example, the DAYS_EMPLOYED column indicates that some individuals have been employed for 1,000 years, which is clearly an anomaly. We handle it by dropping the outliers only.

HANDLING CATEGORICAL VALUES



Some categorical values are ordinal, so we use label encoding for them. Others are not ordinal, so we apply one-hot encoding. However, some columns have many categorical values, so we aggregate them to reduce the number of features after one-hot encoding.

MACHINE LEARNING RESULT

ALGORITHMS AND HYPERPARAMETER TUNING

LOGISTIC REGRESSION	DECISION TREE	RANDOM FOREST	LIGHT GBM	XGBOOST
class weight: balance penalty: l1	No tuning	No tuning	class weight: balance max_depth: 7	No tuning

TOP 3 BEST SCORE

METRICS	LOGISTIC REGRESSION	LIGHT GBM	XGBOOST
ROC AUC	0.75	0.76	0.75
RECALL	0.68	0.03	0.04
PRECISION	0.17	0.53	0.47

From 5 algorithms, Logistic Regression, Light GBM, and XGBoost perform a good score but we only use hyperparameter tuning on Logit and Light GBM due to the limitation of computer's performance in running the model.

We will use the Light GBM algorithm model because it achieved the highest ROC AUC score of 0.76. Given the extreme class imbalance, we need the model to effectively distinguish between the positive and negative classes.

BUSINESS RECOMMENDATION

HOME CREDIT

Recommendation	Focus on Risk Mitigation and Customer Education	Enhance Credit Evaluation and Support	Educational and Community Engagement Initiatives
Insights	High default risk among blue-collar workers with secondary education, primarily in rural/sub-urban areas.	Lower educational attainment correlates with higher default risk.	Professional and higher education groups have lower default rates, vice versa.
Actionable Items	1. Tailored Loan Products: Develop specialized loan products with stricter terms for high-risk occupations, ensuring clear understanding of repayment obligations.	3. Enhanced Credit Scoring Models: Incorporate occupation and education level into credit scoring to better predict default risk and adjust lending criteria accordingly.	5. Financial Literacy Programs: Implement targeted financial literacy workshops and resources for borrowers with secondary education to improve financial management skills.
	2. Risk-Based Pricing: Offer differential interest rates based on risk assessment, providing incentives for timely repayments.	4. Close Monitoring and Support: Establish monitoring systems for early detection of repayment difficulties and provide support or restructuring options for at-risk borrowers.	6. Community Outreach: Engage with local communities to build trust and provide education on financial products and responsible borrowing.

Aldi Vibriani

THANK YOU

Phone :
6282280471417

Github :
Github.com/AldiVibriani

Email :
Aldivibriani@gmail.com

Linkedin:
Aldi Vibriani