

RAKAMIN ACADEMY VIRTUAL INTERNSHIP

Home Credit Scorecard Model

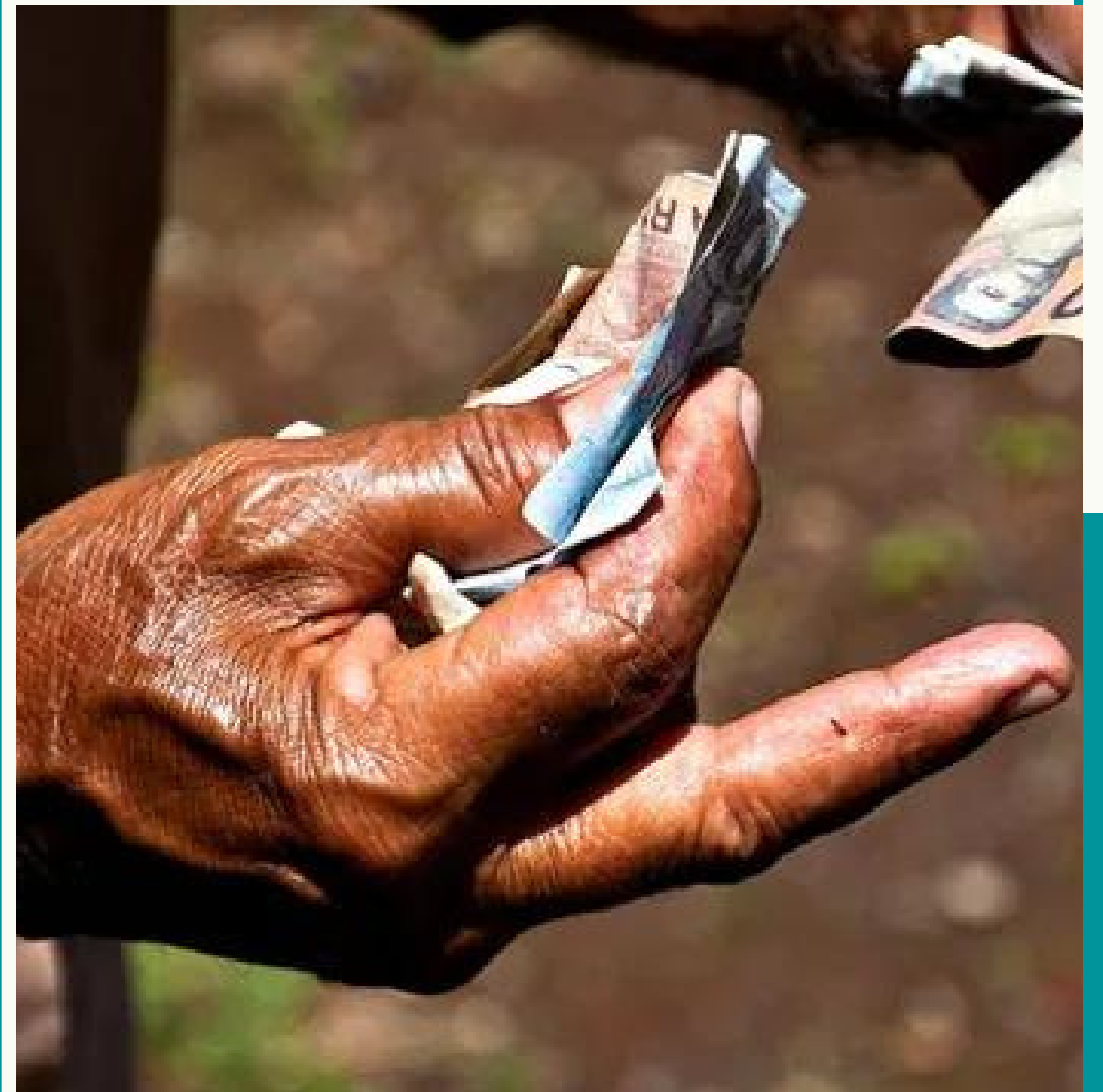
A project by Ahmad Ghulam Sholih

BACKGROUND

Home Credit is focused on promoting financial inclusion for unbanked population, safeguarding them from bad lenders. They call upon data scientists to optimize their data using statistical methods and machine learning. The objective is to avoid denying credit to qualified clients and offer loans with terms conducive to successful repayment.

OBJECTIVE

Develop a robust machine learning model with the capability to predict customers' creditworthiness and minimizing rejections for those with the ability to repay.



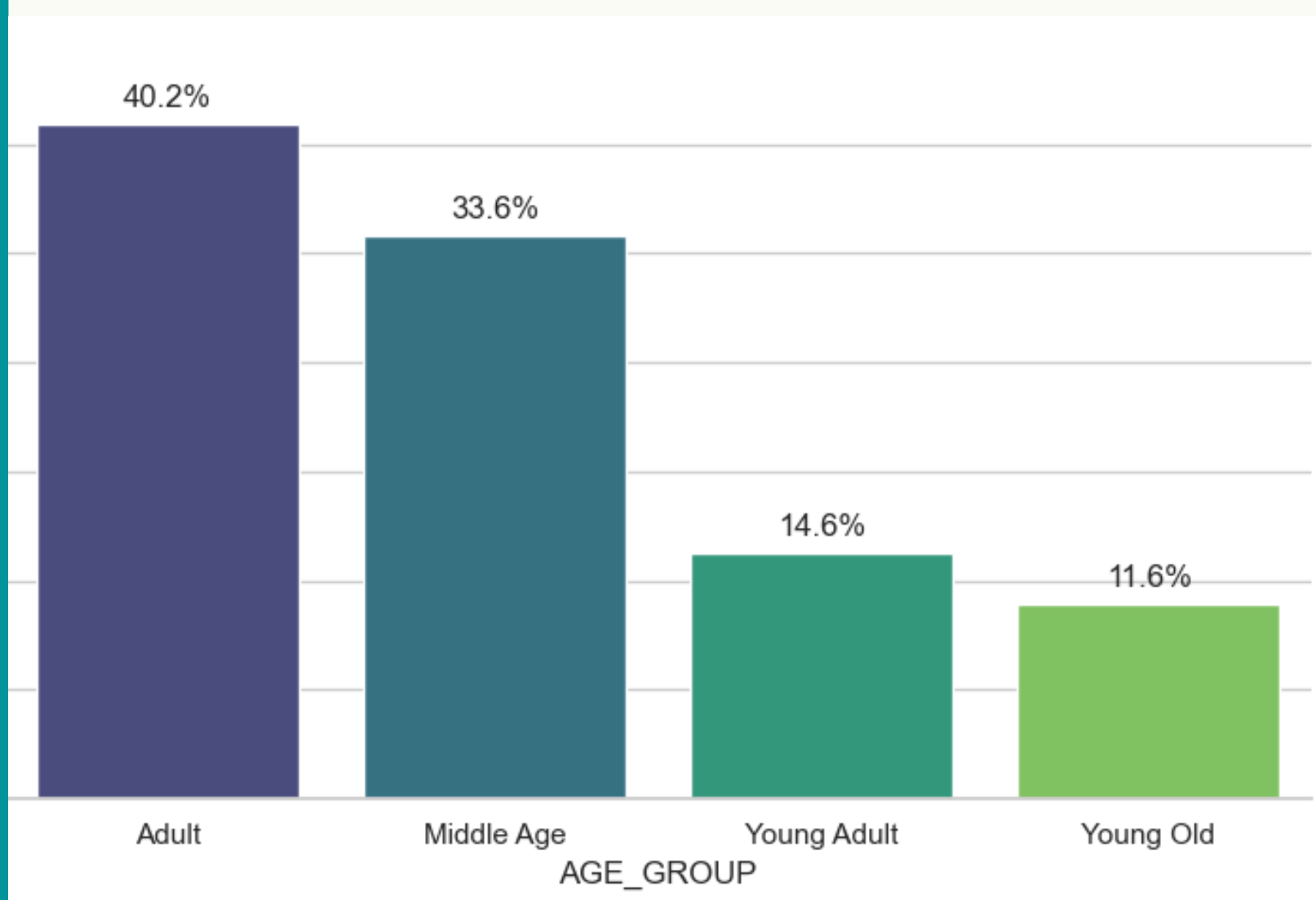
GLIMPSE OF THE DATA

SOURCE	Kaggle
SHAPE	<ul style="list-style-type: none">• Rows : 307511• Columns : 122
MISSING VALUES	<ul style="list-style-type: none">• Delete columns with > 60% NA• Fill categorical with mode, numerical with median
OUTLIERS	<ul style="list-style-type: none">• Delete extreme outliers in 'DAYS_EMPLOYED' and 'AMT_INCOME_TOTAL'

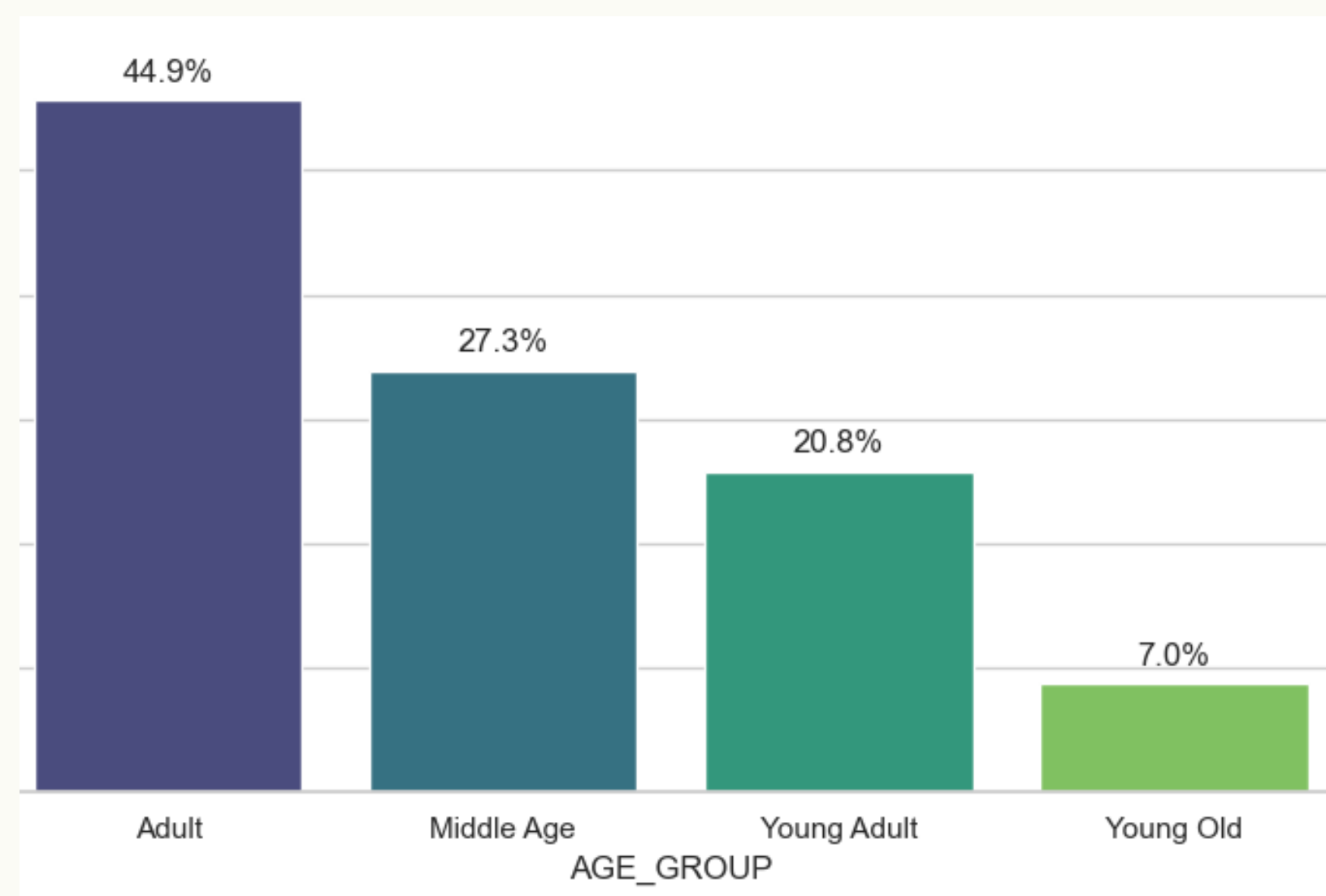
DISTRIBUTION BY AGE GROUP

The young adult cohort (age ≤ 30 years) faces the highest payment difficulties, with a relative increase of 6.2% concerning the number of contracts made by this group. In contrast, the young old demographic (age > 60 years) demonstrates the lowest payment difficulties, reflecting a decrease of 4.6% relative to their contract volume. Notably, adults aged 31 to 45 constitute the most frequent borrowers, accounting for 40.2% of the total contracts.

DATA INSIGHT



All Data

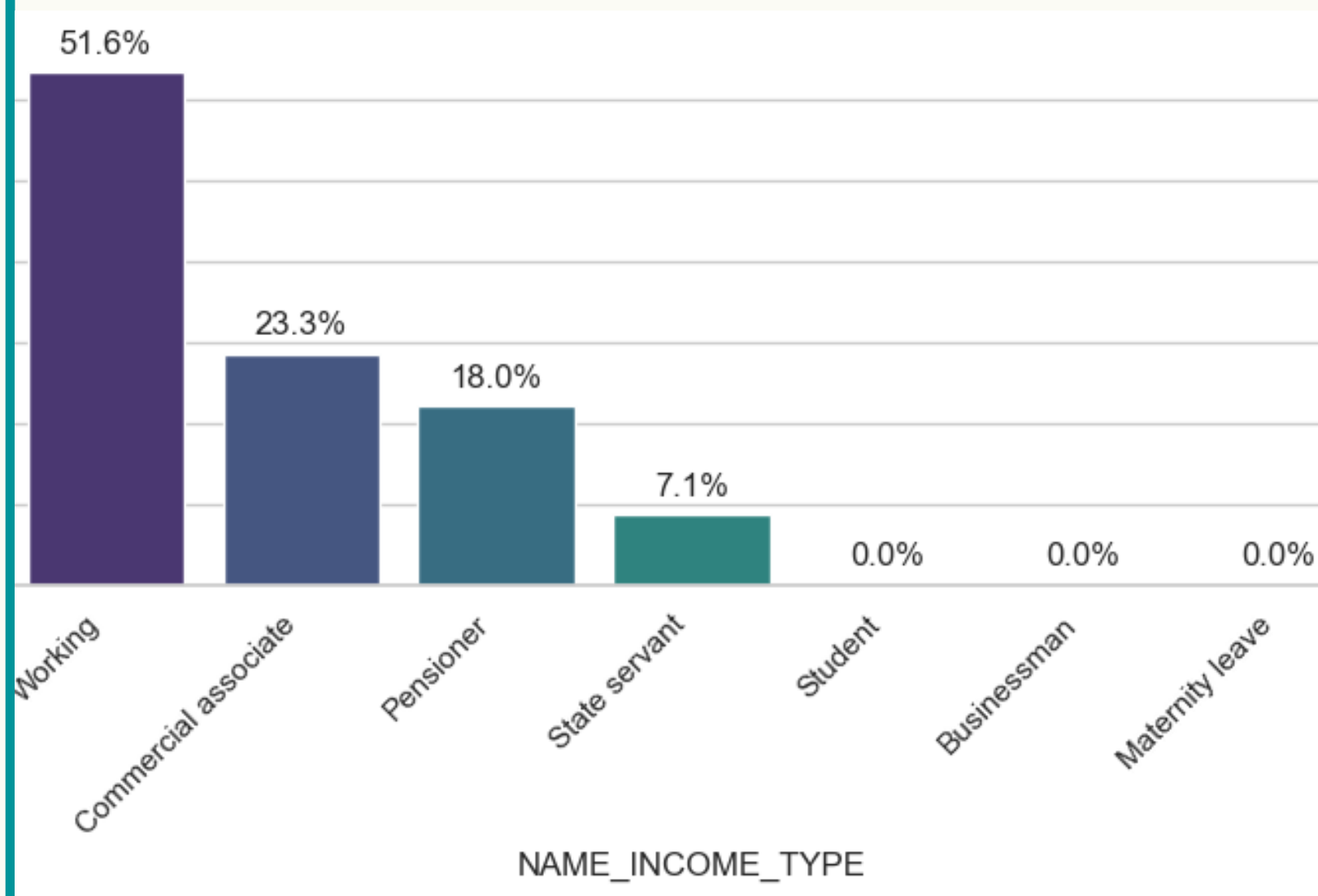


Defaulted

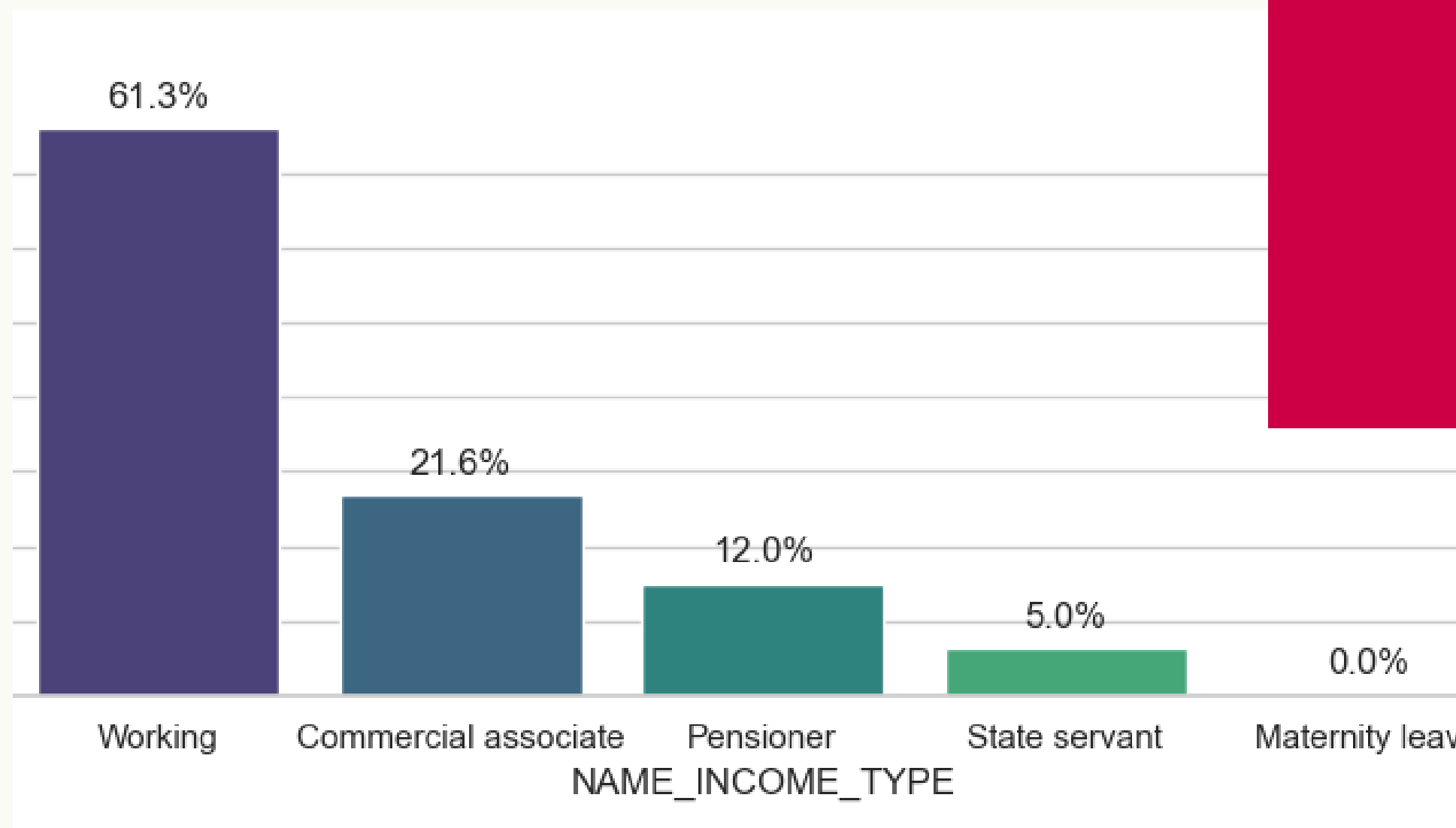
DISTRIBUTION BY INCOME TYPE

The working income type is the dominant category among total borrowers, comprising 51.6%, but it also faces the highest payment problems with a relative increase of 9.7% concerning the total contracts made by this group. In contrast, the pensioner income type exhibits the lowest difficulty in payment, showing a percentage decrease of 6% relative to the total contracts made by pensioners. Meanwhile, the proportions of student, businessman, and maternity income types are minimal,

DATA INSIGHT



All Data



Defaulted

Data Preprocessing

Feature Extraction

CREDIT_GOOD_PRICE_RATIO
EXTERNAL_SOURCE_AVG
CONTACT_INFO_MISMATCH
SOCIAL_SURROUNDING_DEF

Feature Encoding

OHE used for columns with > 2 unique values.

Label encoding used for columns with only 2 values

Data Alignment

Align train and test data

Feature Scaling

StandardScaler

Data Preprocessing

Feature Selection

22 highest correlated features with target

Imbalance Handling

SMOTE after data split to prevent data leak

MODELING

MODEL COMPARISON



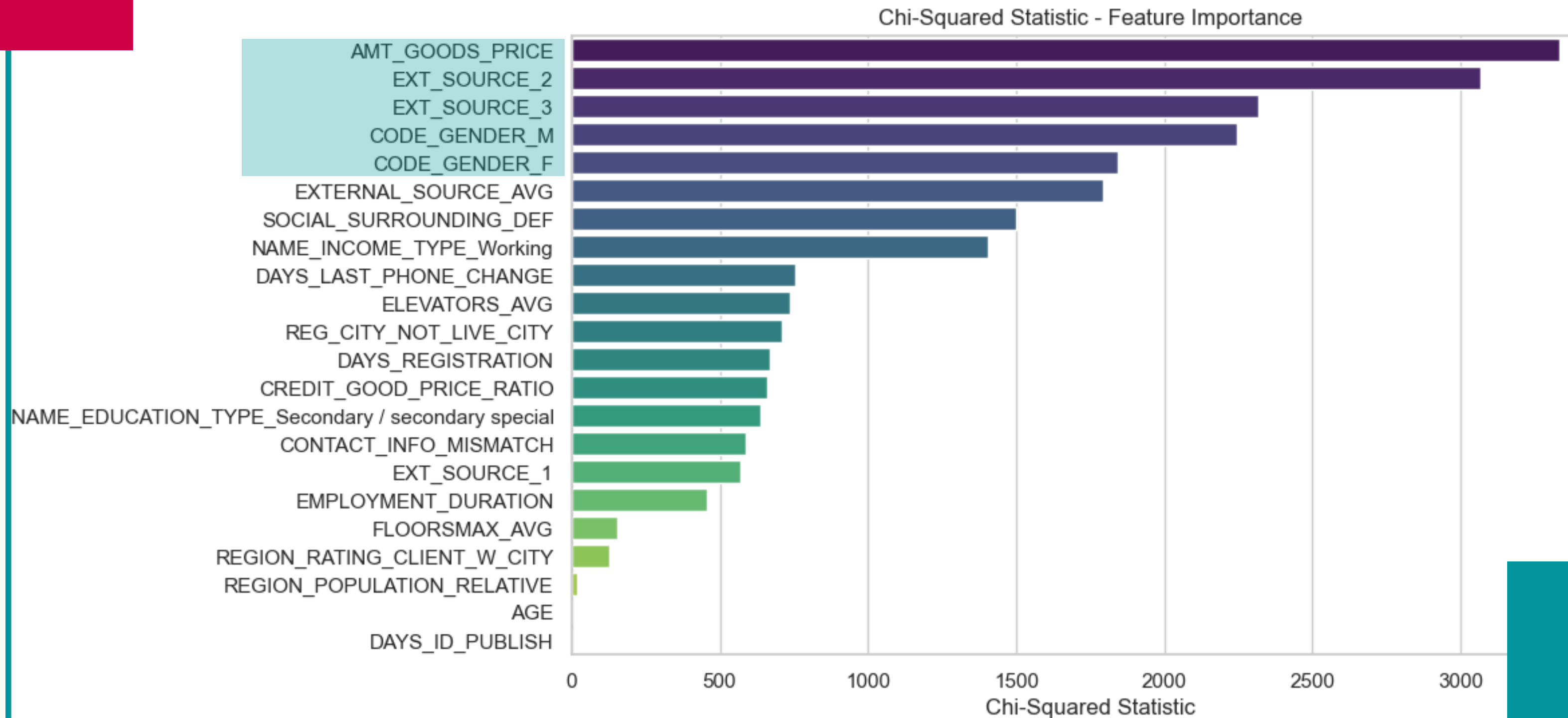
BEST MODEL
LOGISTIC REGRESSION

0.7325

ROC AUC SCORE

No further improvement after
hyperparameter tuning

Feature Importance



BUSINESS RECOMMENDATIONS

DYNAMIC AND PERSONALIZED CREDIT TERMS



- Implement dynamic credit limit adjustments
- Tailor credit terms, repayment schedules, and interest rates

FINANCIAL LITERACY INITIATIVES



- Launch credit education programs
- Provide resources and tools for budgeting, financial planning, and debt management

CUSTOMER OUTREACH PROGRAMS



- Targeted marketing and product development
- Collaborate with educational institutions and offer specialized financial solutions

RISK MITIGATION STRATEGIES



- Develop comprehensive risk mitigation strategies considering various factors
- Implement real-time monitoring and alerts for potential repayment challenges

FOR FULL PROJECT DOCUMENTATION,
PLEASE VISIT

MY GITHUB REPOSITORY



THANK YOU!

Looking forward to working
together.