CREDIT CARD APPROVAL SYSTEM

THE METRICS MARVINS

RONALD NYAGAKA

MAUREEN KITANG'A

CLIFF SHITOTE

BEATRICE KIRUI

PAUL KAMAU

LEAH KATIWA

BLUEPRINT

This project was carried out using the CRISP-DM methodology

BUSINESS UNDERSTANDING

DATA UNDERSTANDING

DATA PREPARATION

MODELING

EVALUATION

DEPLOYMENT

OVERVIEW

Credit score cards are widely used in the financial industry to assess the risk of issuing credit cards. Traditional approaches rely on financial metrics but can be influenced by economic changes.

However, these methods often lack transparency, making it difficult to explain acceptance or rejection decisions.

PROBLEM STATEMENT

In the realm of credit scoring, traditional methods based on financial metrics face limitations in accurately predicting creditworthiness due to economic fluctuations. The introduction of machine learning algorithms offers potential improvements in accuracy, yet their lack of transparency in explaining acceptance or rejection decisions poses challenges.

To address these issues, we will create a reliable and transparent credit scoring system that assesses the risk of issuing credit cards based on applicants' personal information, enabling informed decision-making by financial institutions and ensuring fair treatment of applicants.

BUSINESS OBJECTIVES

MAIN OBJECTIVE

- Develop a credit scoring model that incorporates personal and social factors and machine learning algorithms to enhance the accuracy of creditworthiness predictions.

SPECIFIC OBJECTIVES

- Improve transparency in credit scoring by utilizing interpretable machine learning techniques, allowing for clear explanations of acceptance or rejection decisions to customers and regulatory bodies.
- Mitigate the impact of economic fluctuations on credit scoring models by incorporating dynamic factors and adapting the model to change

DATA UNDERSTANDING

APPLICATION.CSV
(438,557)

ID
CODE_GENDER
FLAG_OWN_CAR
FLAG_OWN_REALTY
CNT_CHILDREN
AMT_INCOME_TOTAL
NAME_INCOME_TYPE
NAME_EDUCATION_TYPE
NAME_FAMILY_STATUS
NAME_HOUSING_TYPE
DAYS_BIRTH
DAYS_EMPLOYED
FLAG_MOBIL
FLAG_WORK_PHONE
FLAG_PHONE
FLAG_EMAIL
OCCUPATION_TYPE
CNT_FAM_MEMBERS

CREDITS.CSV
(1,048,575)

ID
MONTHS_BALANCE
STATUS

MERGED DATASET
(777,715)

ID
CODE_GENDER
FLAG_OWN_CAR
FLAG_OWN_REALTY
CNT_CHILDREN
AMT_INCOME_TOTAL
NAME_INCOME_TYPE
NAME_EDUCATION_TYPE
NAME_FAMILY_STATUS
NAME_HOUSING_TYPE
DAYS_BIRTH
DAYS_EMPLOYED
FLAG_MOBIL
FLAG_WORK_PHONE
FLAG_PHONE
FLAG_EMAIL
OCCUPATION_TYPE
CNT_FAM_MEMBERS
MONTHS_BALANCE
STATUS

DATA PREPARATION

THE DATASET WAS PREPARED FOR MODELING THROUGH THE FOLLOWING STEPS

CLEANING THE DATA

TO RESOLVE ERRONEOUS VALUES

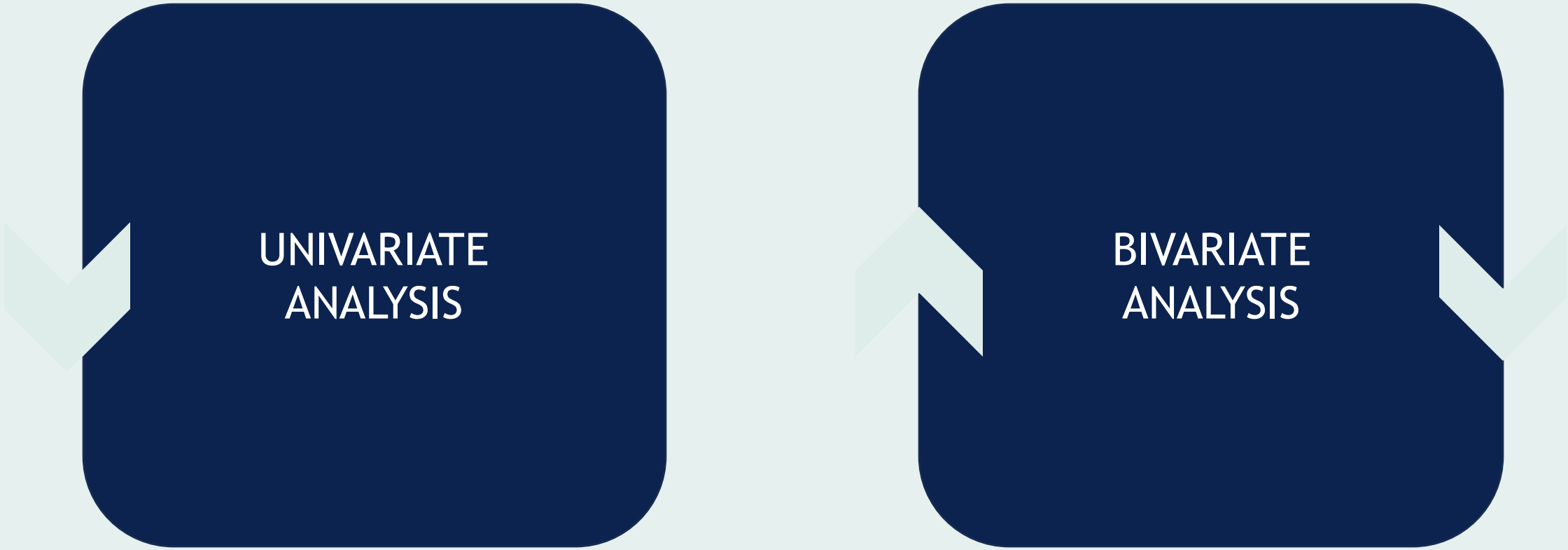
PERFORMING EDA

TO UNVEIL THE UNDERLYING DISTRIBUTIONS WITHIN THE DATASET

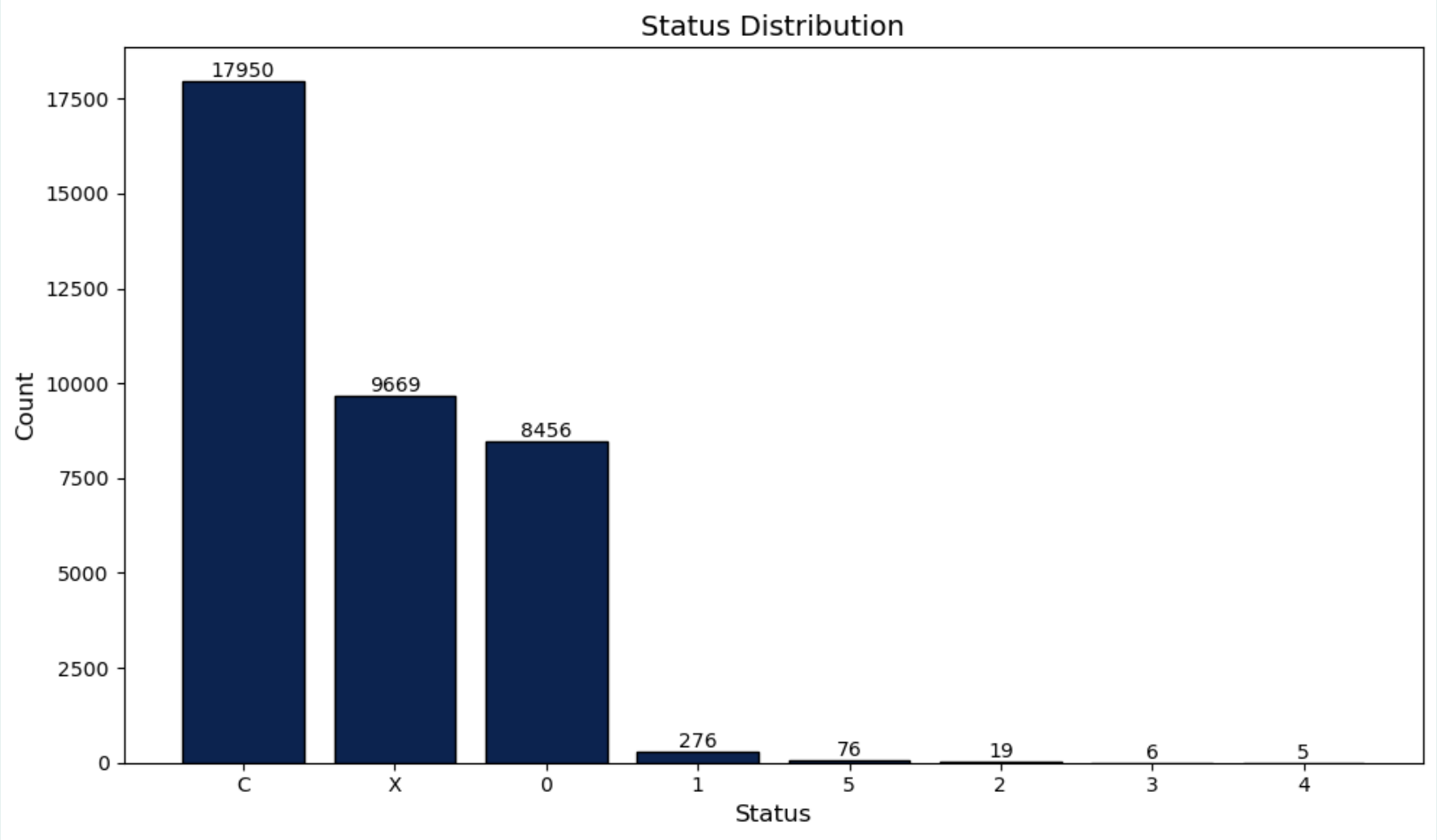
DATA PREPROCESSING

TO COME UP WITH NEW FEATURES AND DO FINAL FEATURE SELECTION

EXPLORATORY DATA ANALYSIS

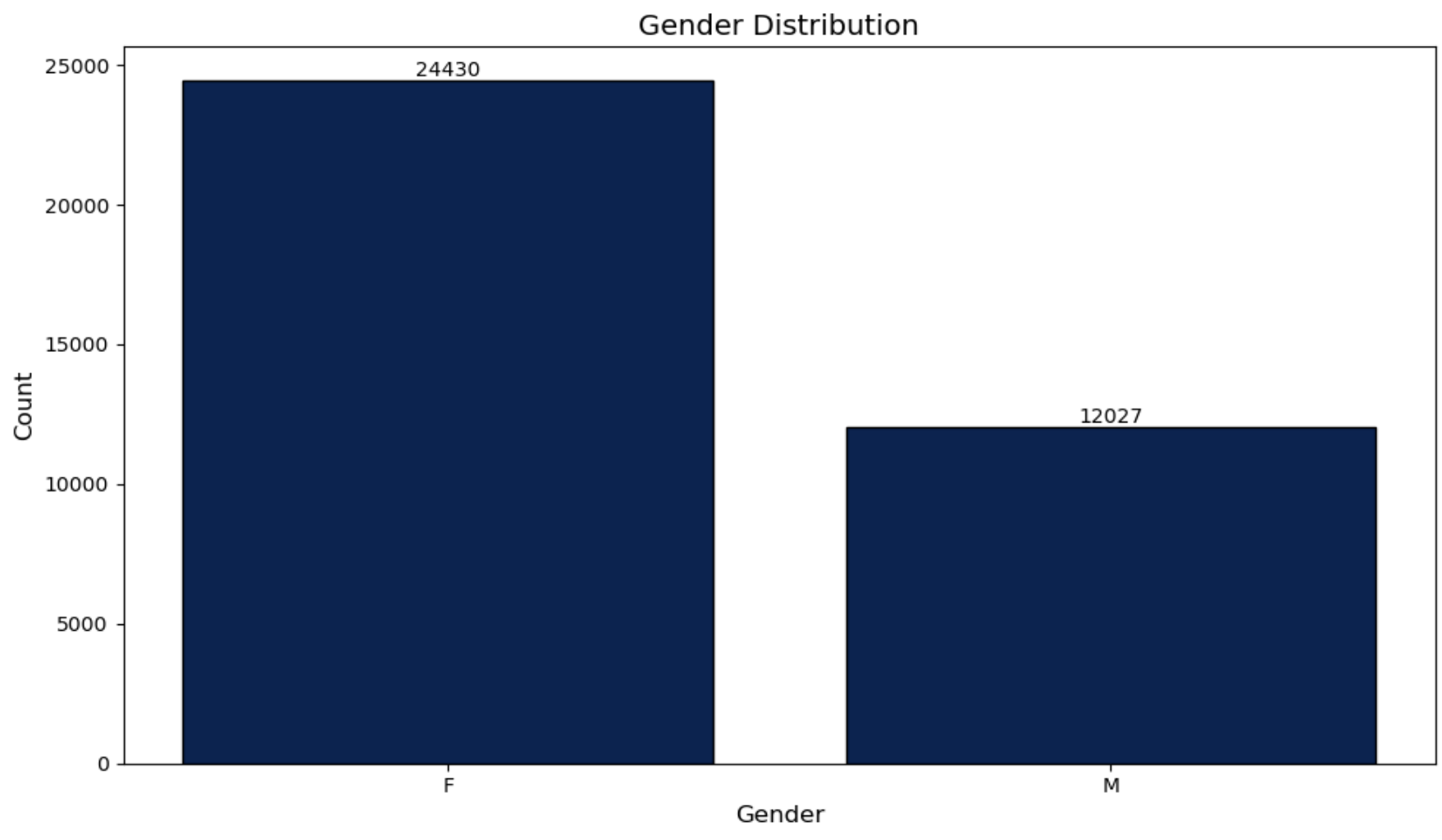


UNIVARIATE ANALYSIS OF STATUS



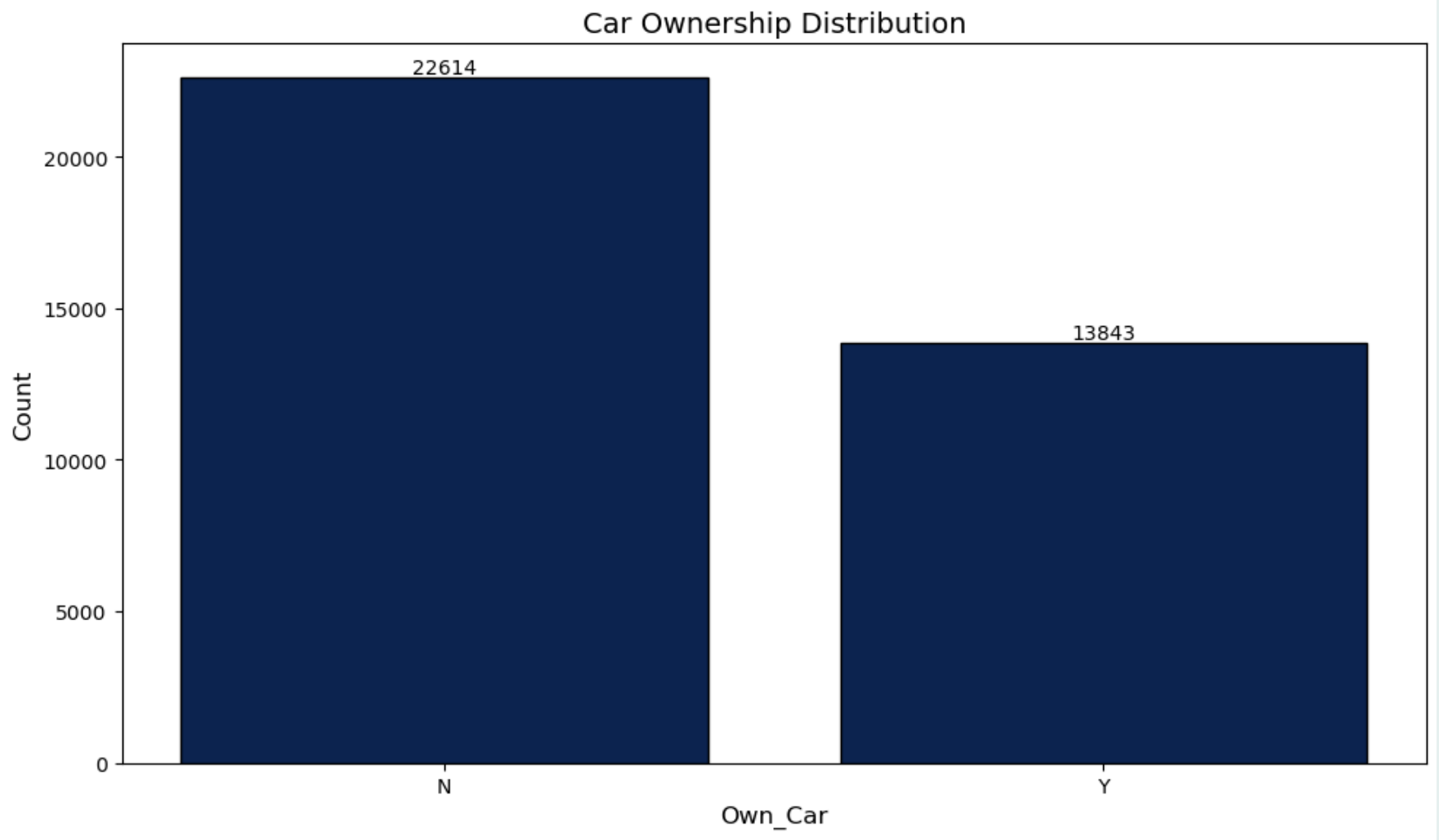
THE MAJORITY OF APPLICANTS FALL UNDER STATUS '**C**' AND '**X**'

UNIVARIATE ANALYSIS OF GENDER



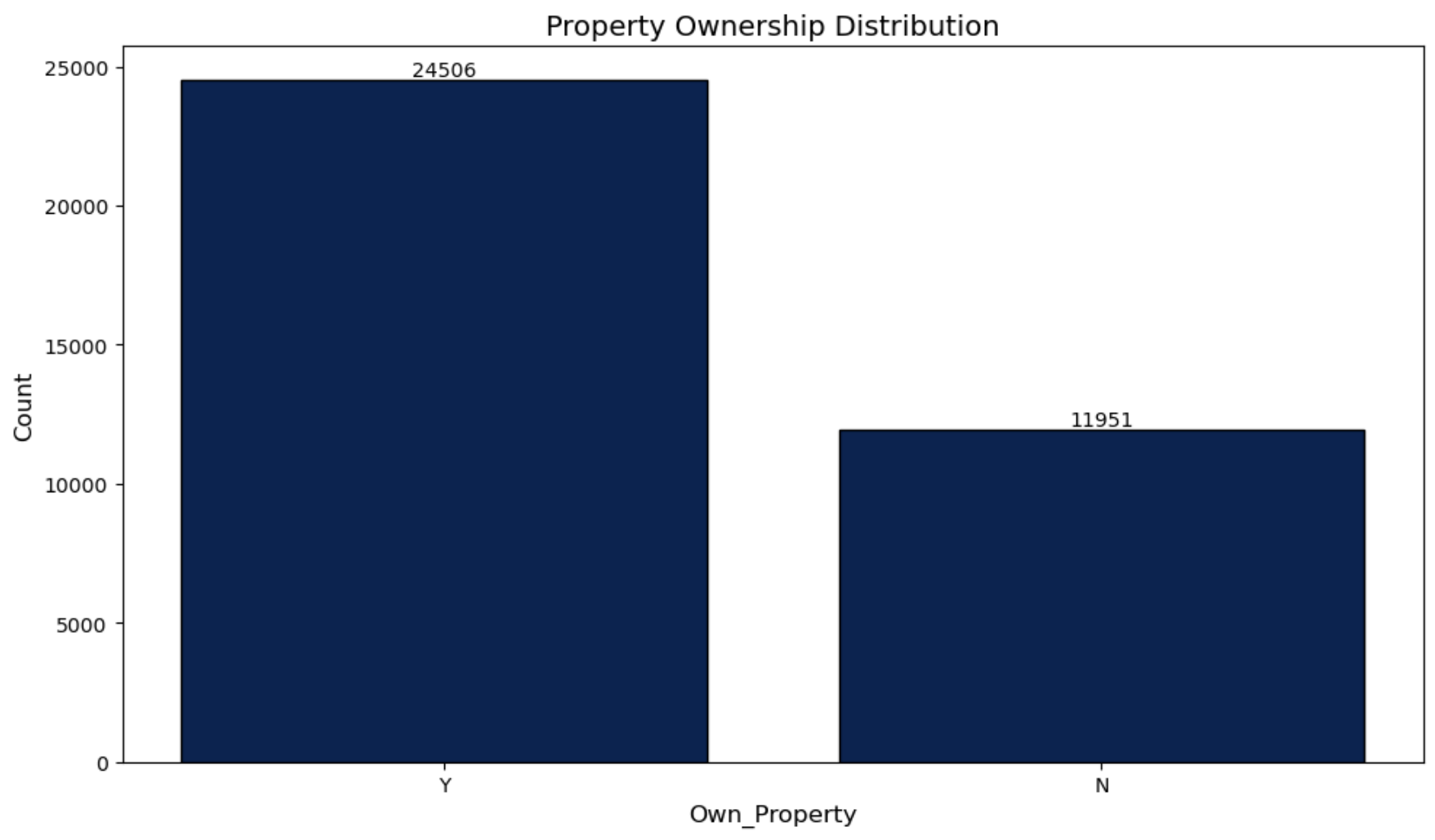
THE MAJORITY OF APPLICANTS ARE **FEMALES** WHICH ACCOUNTS FOR **67%** OF TOTAL APPLICANTS

UNIVARIATE ANALYSIS OF CAR OWNERSHIP



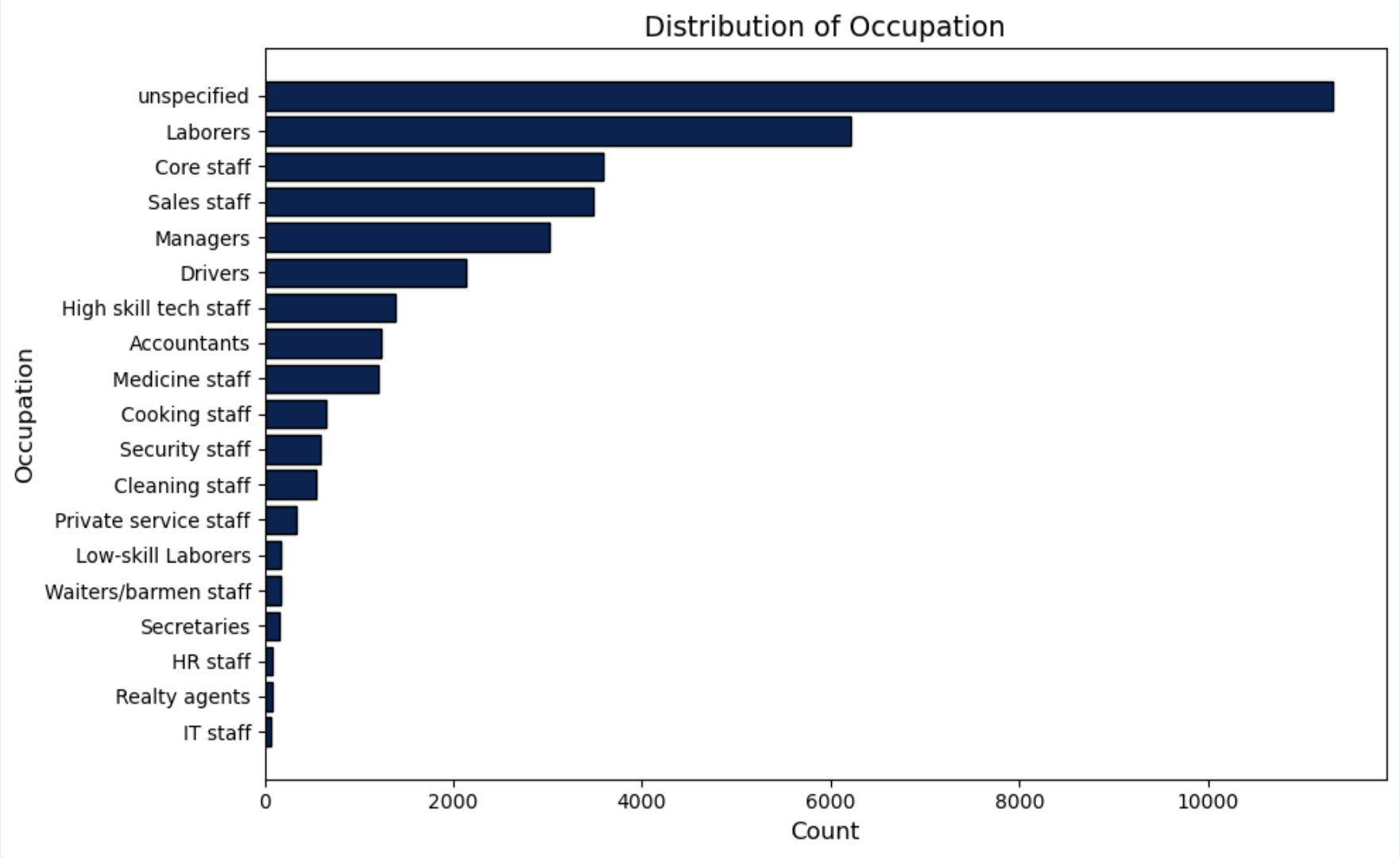
THE MAJORITY OF APPLICANTS, ACCOUNTING FOR **62%**, OWN A CAR.

UNIVARIATE ANALYSIS OF PROPERTY OWNERSHIP



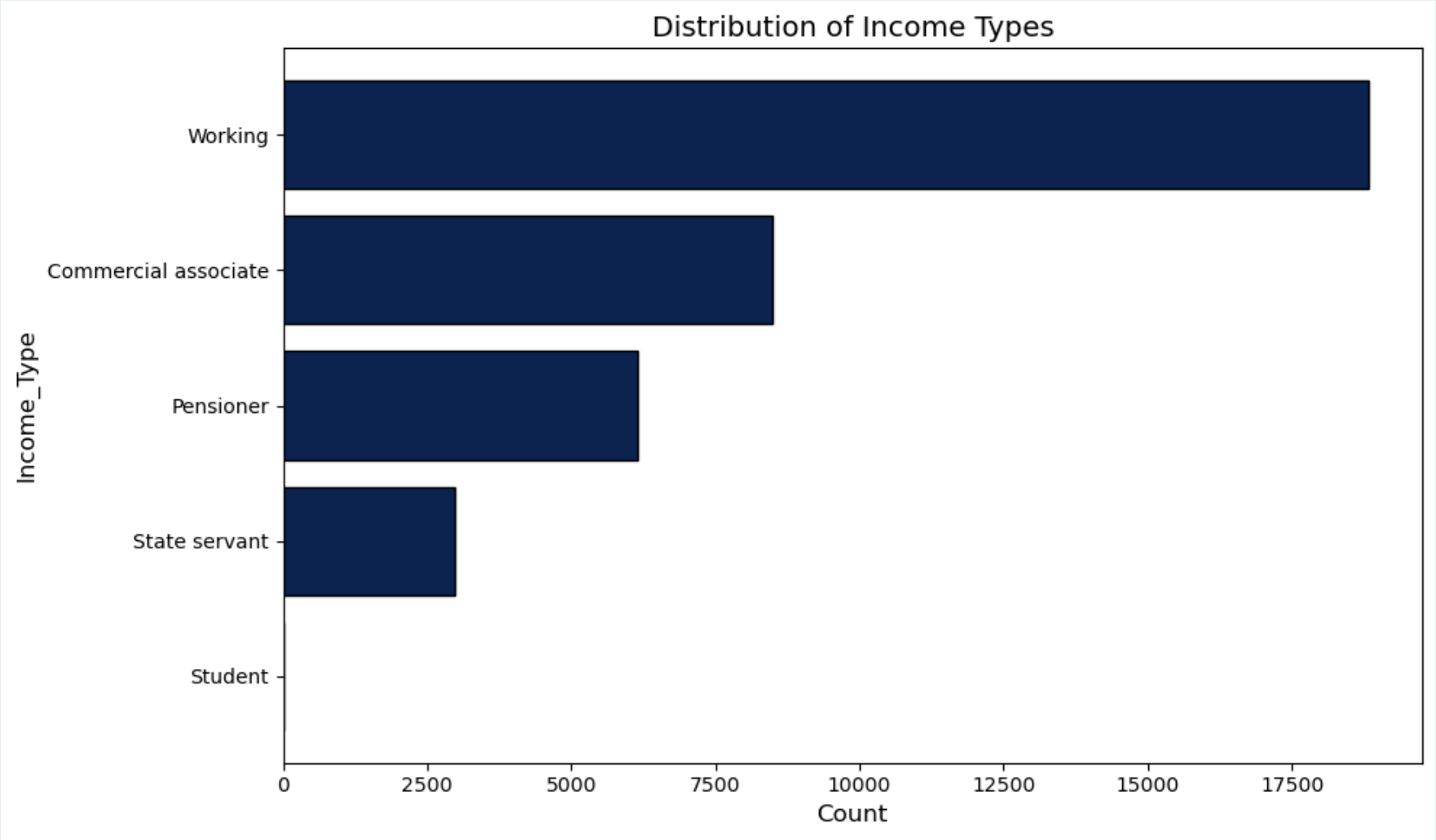
THE MAJORITY OF APPLICANTS, ACCOUNTING FOR **67%**, OWN A PROPERTY.

UNIVARIATE ANALYSIS OF OCCUPATION



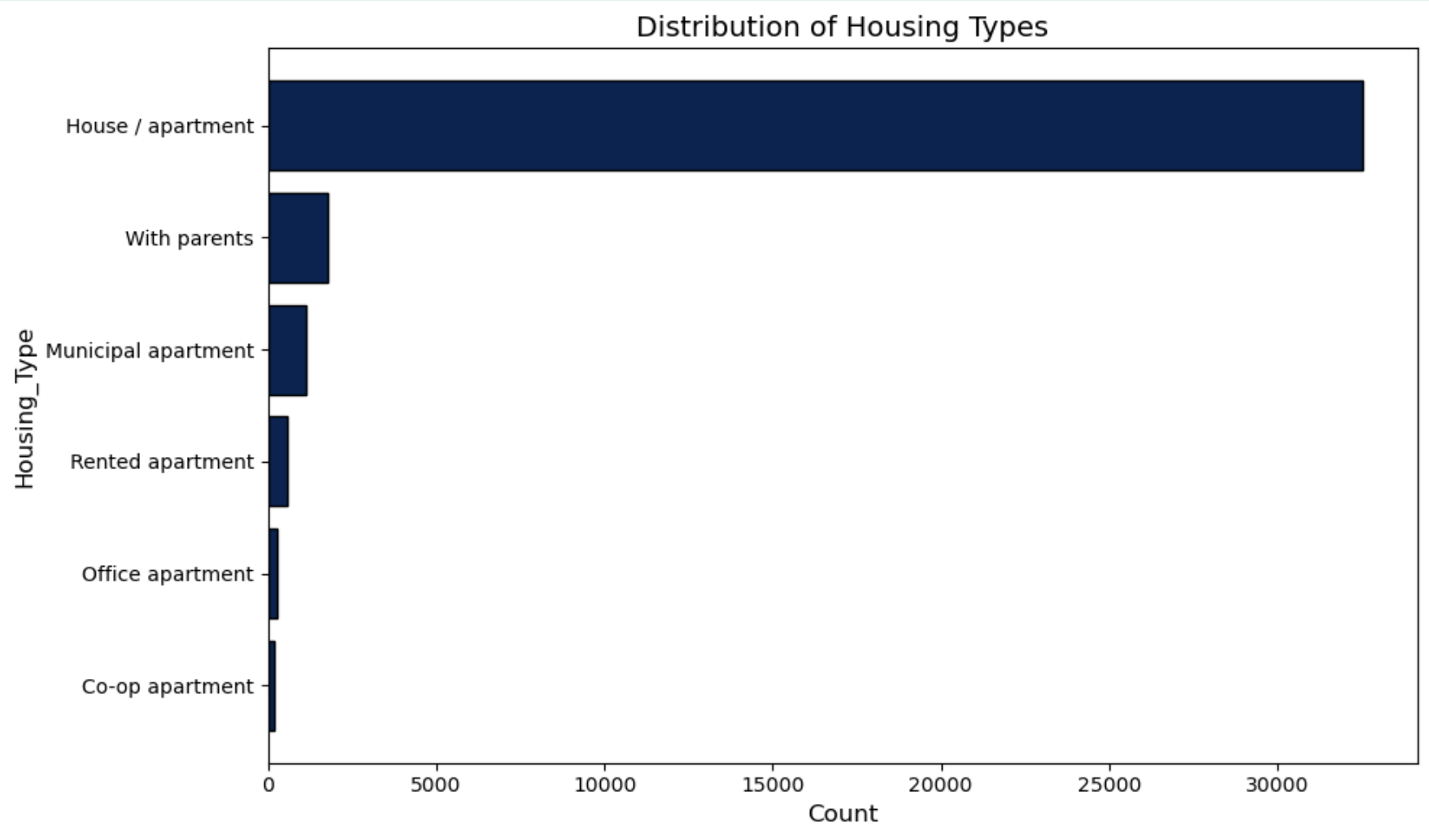
MOST OF THE APPLICANTS DID NOT SPECIFY THEIR OCCUPATIONS, THE OTHER TOP TWO CATEGORIES WERE; **LABORERS** AND **CORE STAFF.**

UNIVARIATE ANALYSIS OF INCOME TYPES



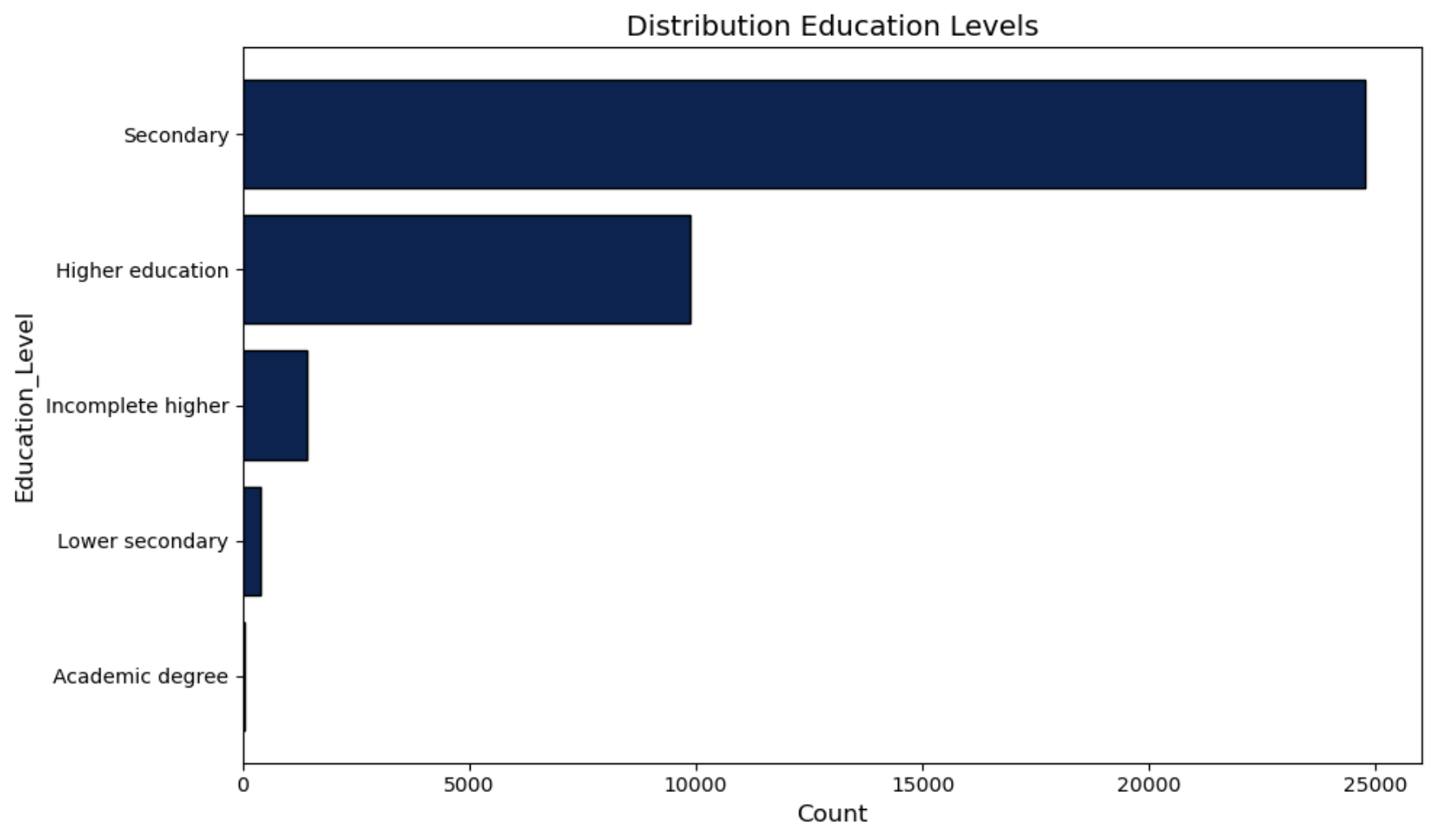
`**WORKING**` STANDS OUT AS THE MOST PREVALENT INCOME TYPE.

UNIVARIATE ANALYSIS OF HOUSING TYPES



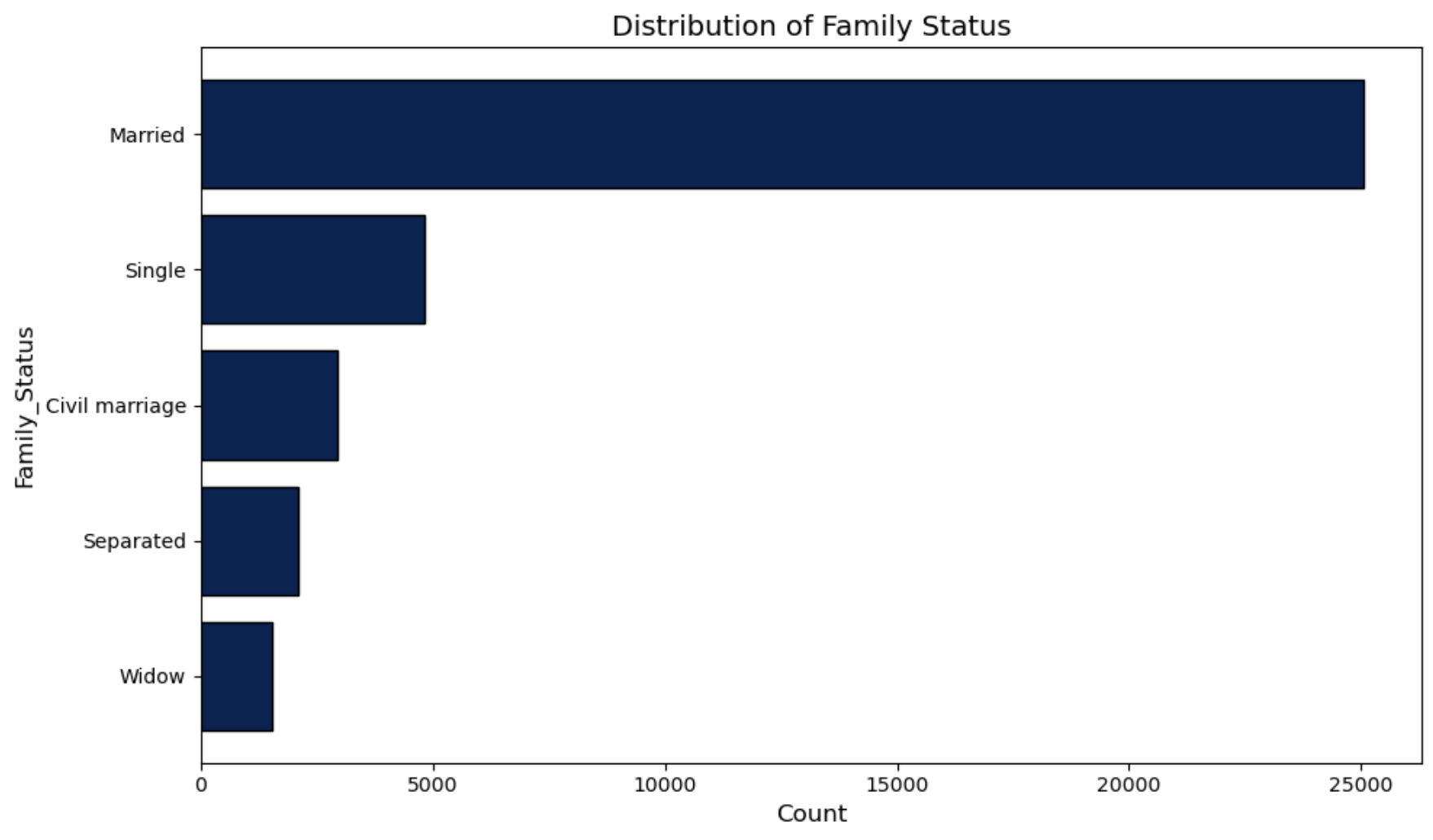
MOST OF THE APPLICANTS LIVE IN A **HOUSE/APARTMENT** WITH ACCOUNTS FOR **91%** OF THE TOTAL APPLICANTS

UNIVARIATE ANALYSIS OF EDUCATION LEVELS



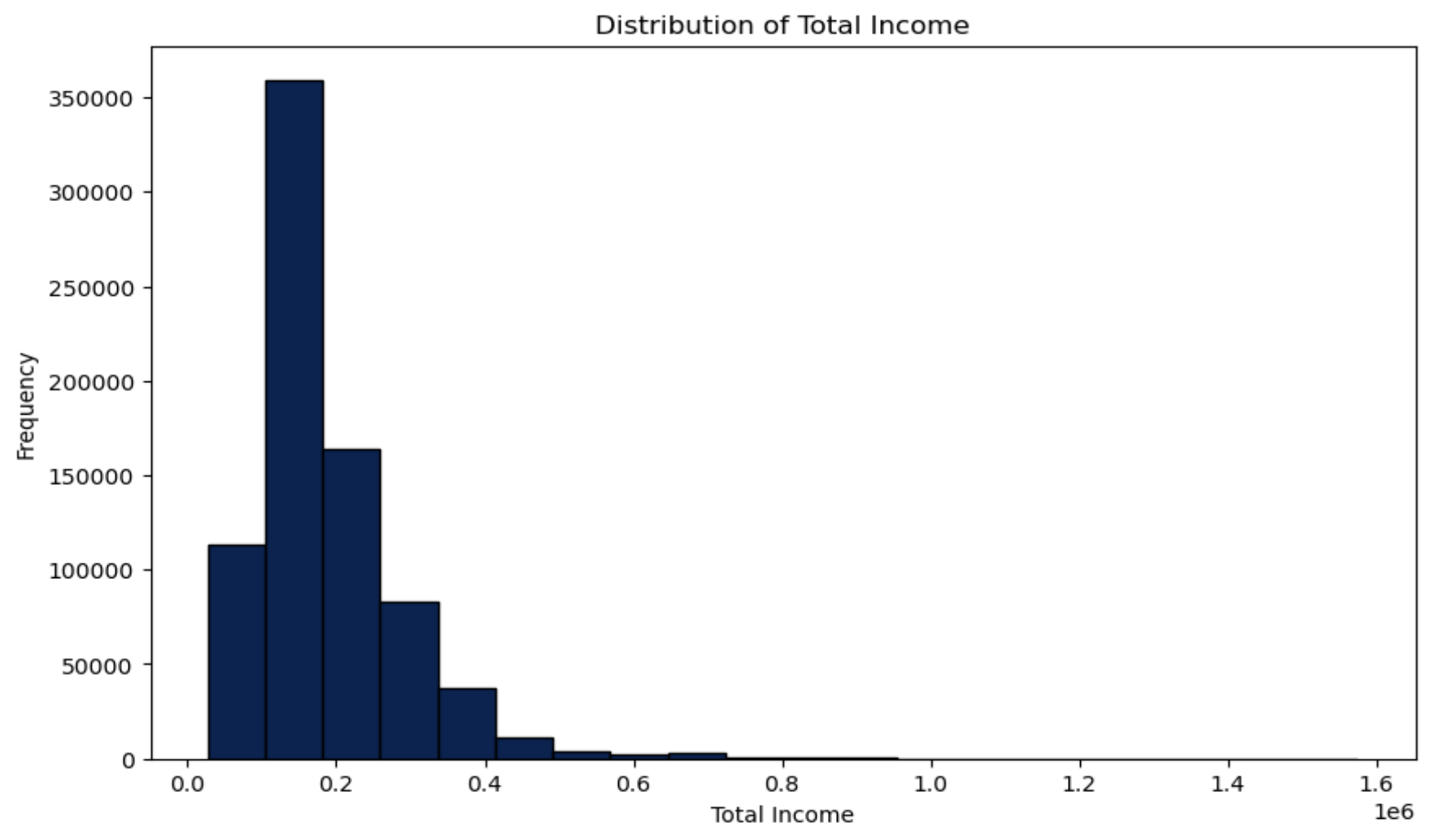
SECONDARY IS THE MOST COMMON EDUCATION LEVEL(66.2%) FOLLOWED BY **HIGHER EDUCATION**(26.3%)

UNIVARIATE ANALYSIS OF FAMILY STATUS



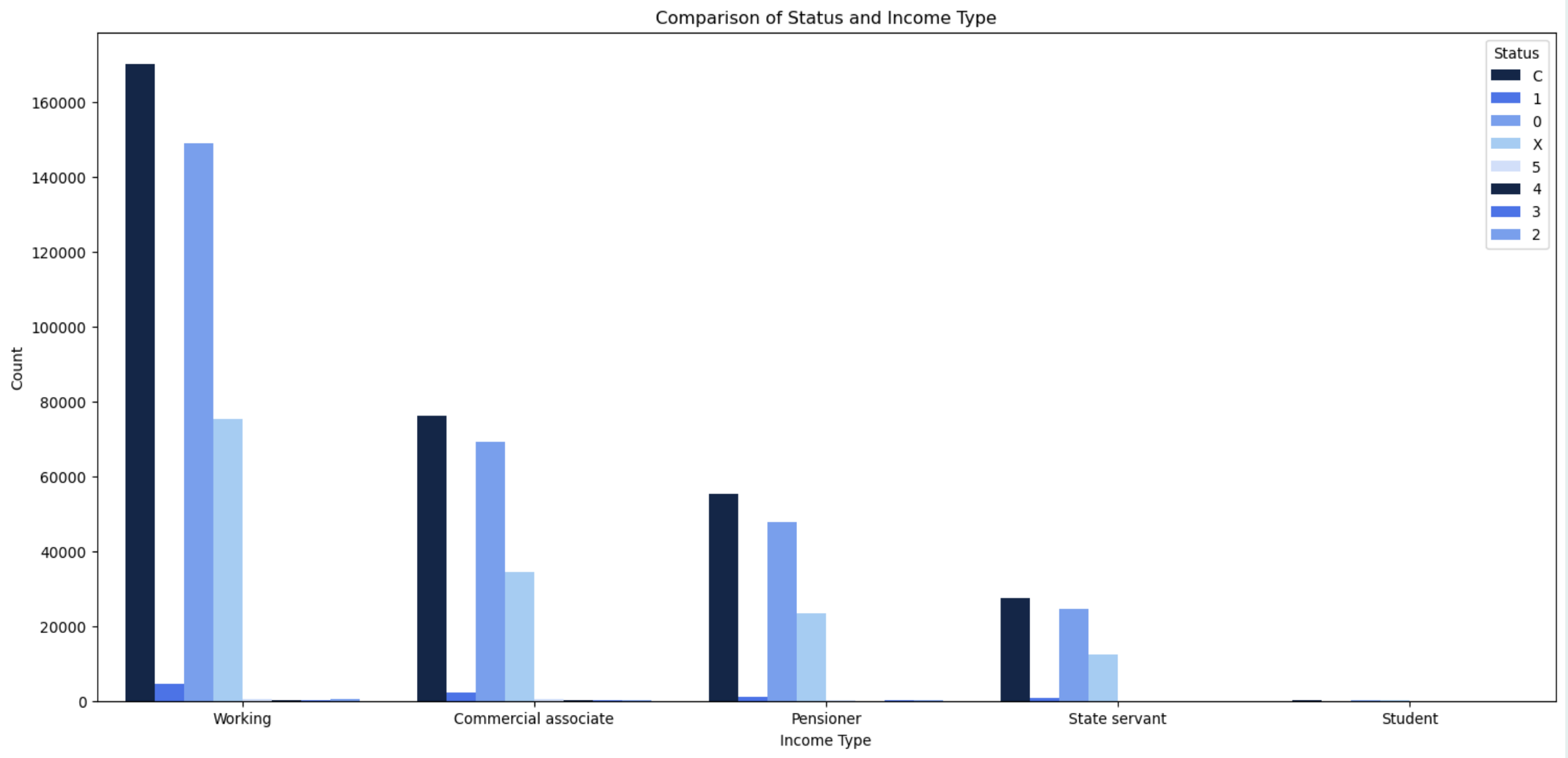
MOST OF THE APPLICANTS ARE **MARRIED(MARRIED AND CIVIL MARRIAGE)** WHICH ACCOUNTS FOR **79%** OF THE TOTAL APPLICANTS

UNIVARIATE ANALYSIS OF TOTAL INCOME



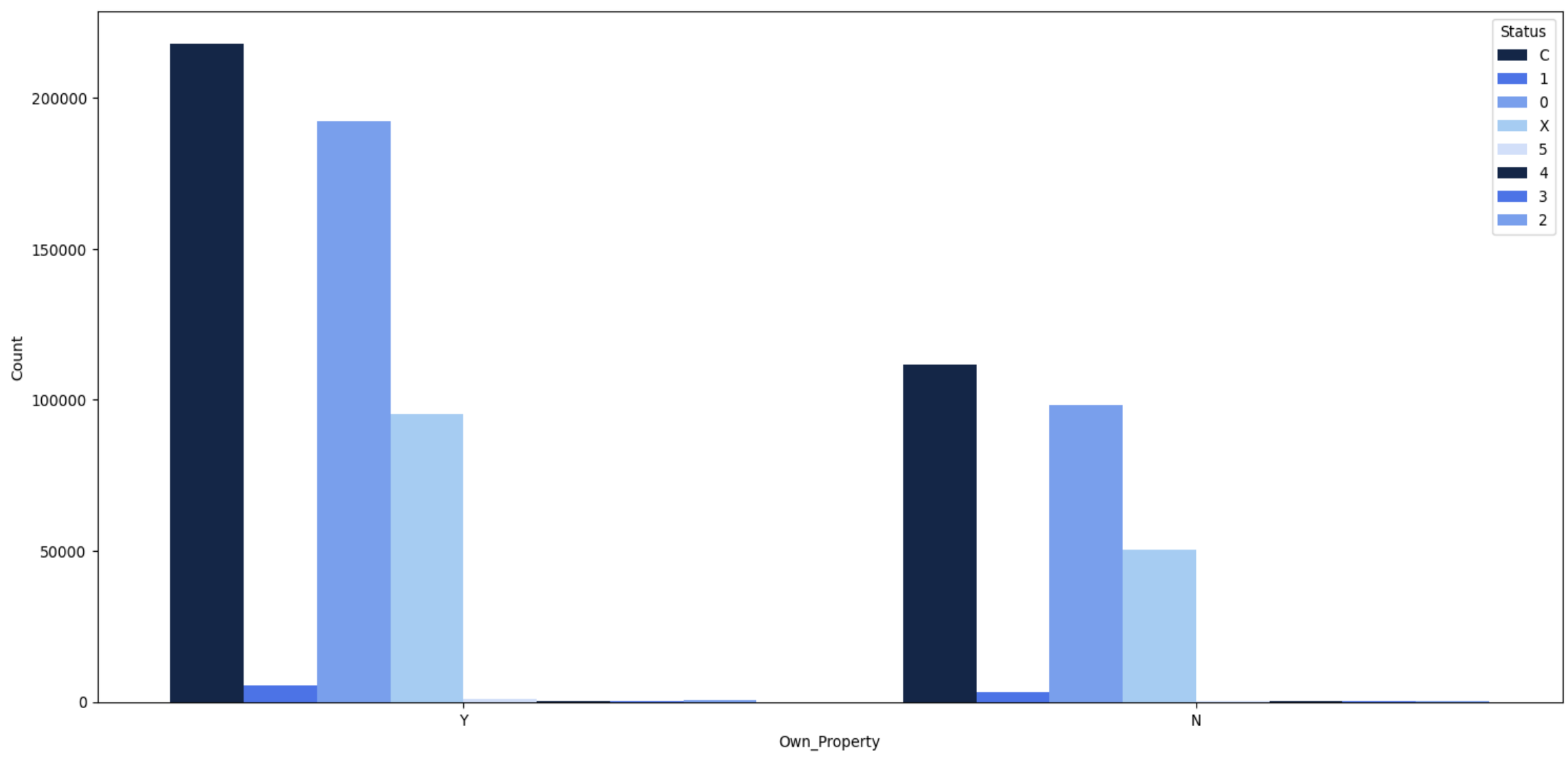
THE MAJORITY OF APPLICANTS EARN BELOW **500,000**, WITH A SIGNIFICANT PORTION FALLING WITHIN THE INCOME RANGE OF **100,000** TO **200,000.**

BIVARIATE ANALYSIS:STATUS VS INCOME TYPE



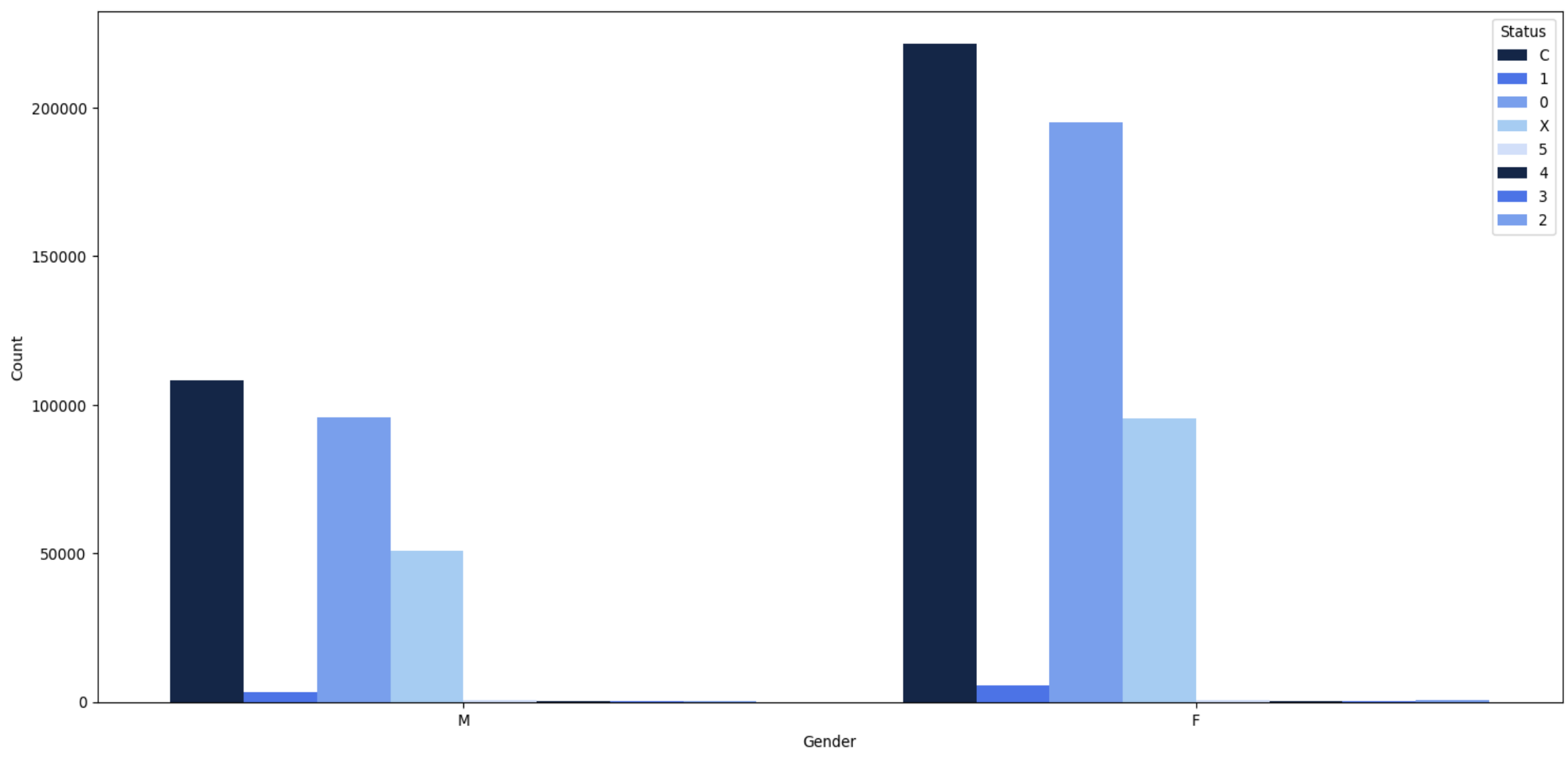
WORKING AND COMMERCIAL ASSOCIATE HAVE HIGHER PROPORTIONS IN ALL LOAN STATUSES

BIVARIATE ANALYSIS:STATUS VS PROPERY OWNERSHIP



MOST OF PROPERTY OWNERS HAVE HIGHEST COUNT IN ALL VARIOUS LOAN STATUSES

BIVARIATE ANALYSIS:STATUS VS GENDER



FEMALES HAVE THE HIGHEST COUNT IN ALL LOAN STATUSES

DATA PREPROCESSING

WE CREATED TWO NEW FEATURES:
STATUS >> CREDIT_APPROVAL_STATUS
TOTAL_INCOME >> INCOME_RANGE

WE DROPPED SOME FEATURES
(ID, NUM_CHILDREN, MOBILE,
RECORDED_DATE(MONTHS),
AGE(DAYS),STATUS, TOTAL_INCOME,
WORK_PHONE,PHONE)

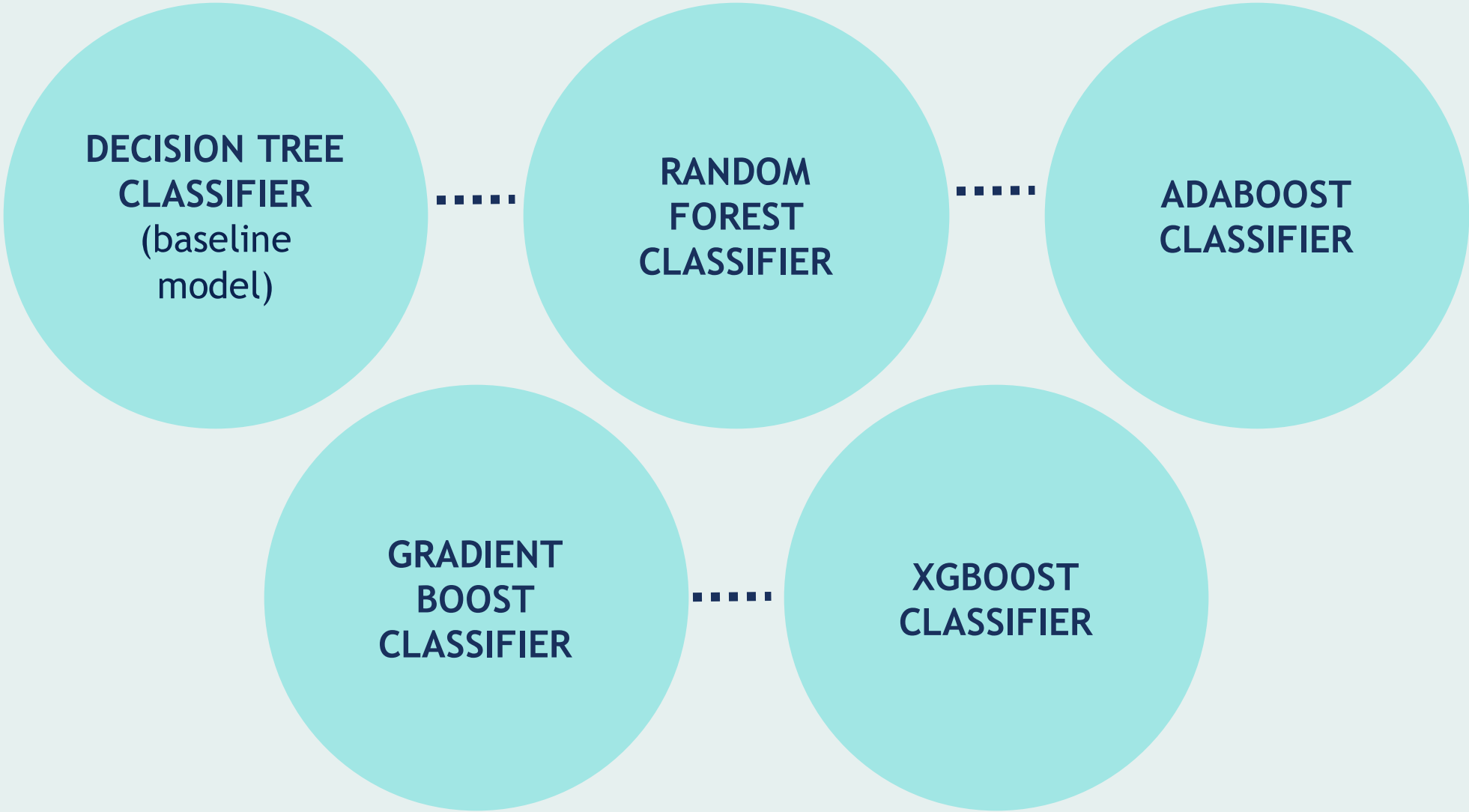
WE USED FEATURE RECURSIVE ELIMINATION TO SELECT THE TOP 12 FEATURES

THE FINAL COLUMNS FOR MODELING:

CREDIT_APPROVAL_STATUS(TARGET)
GENDER
OWN_CAR
OWN_PROPERTY
INCOME_TYPE
INCOME_RANGE
EDUCATION_LEVEL
FAMILY_STATUS
HOUSING_TYPE
EMPLOYMENT_DURATION
OCCUPATION
NUM_FAMILY
AGE

MODELING

WE CHOSE THE FOLLOWING ALGORITHMS FOR OUR MODELING



EVALUATION

Recall score was used to evaluate the performance of the credit card approval system by measuring its ability to correctly identify and approve applicants who are likely to repay their credit card debts.

DECISION TREE CLASSIFIER	0.85
RANDOM FOREST CLASSIFIER	0.84
ADA BOOST CLASSIFIER	0.59
GRADIENT BOOST CLASSIFIER	0.82
XGBOOST CLASSIFIER	0.90

MINIMUM VIABLE PRODUCT

- While XGB Class delivered an impressive recall score, ensuring a higher number of true positive predictions, our incorporation of Decision Tree has further improved the overall prediction accuracy.
- Our Minimum Viable Product for credit card approval prediction combines the power of Decision Tree with the interpretability of SHAP (***SH****APLEY ADDITIVE EXPLANATIONS*). Decision Tree provides accurate predictions, while SHAP helps us understand the influence of each feature in the approval decision. This integration ensures not only precise credit card approval predictions but also transparency and fairness in the decision-making process.

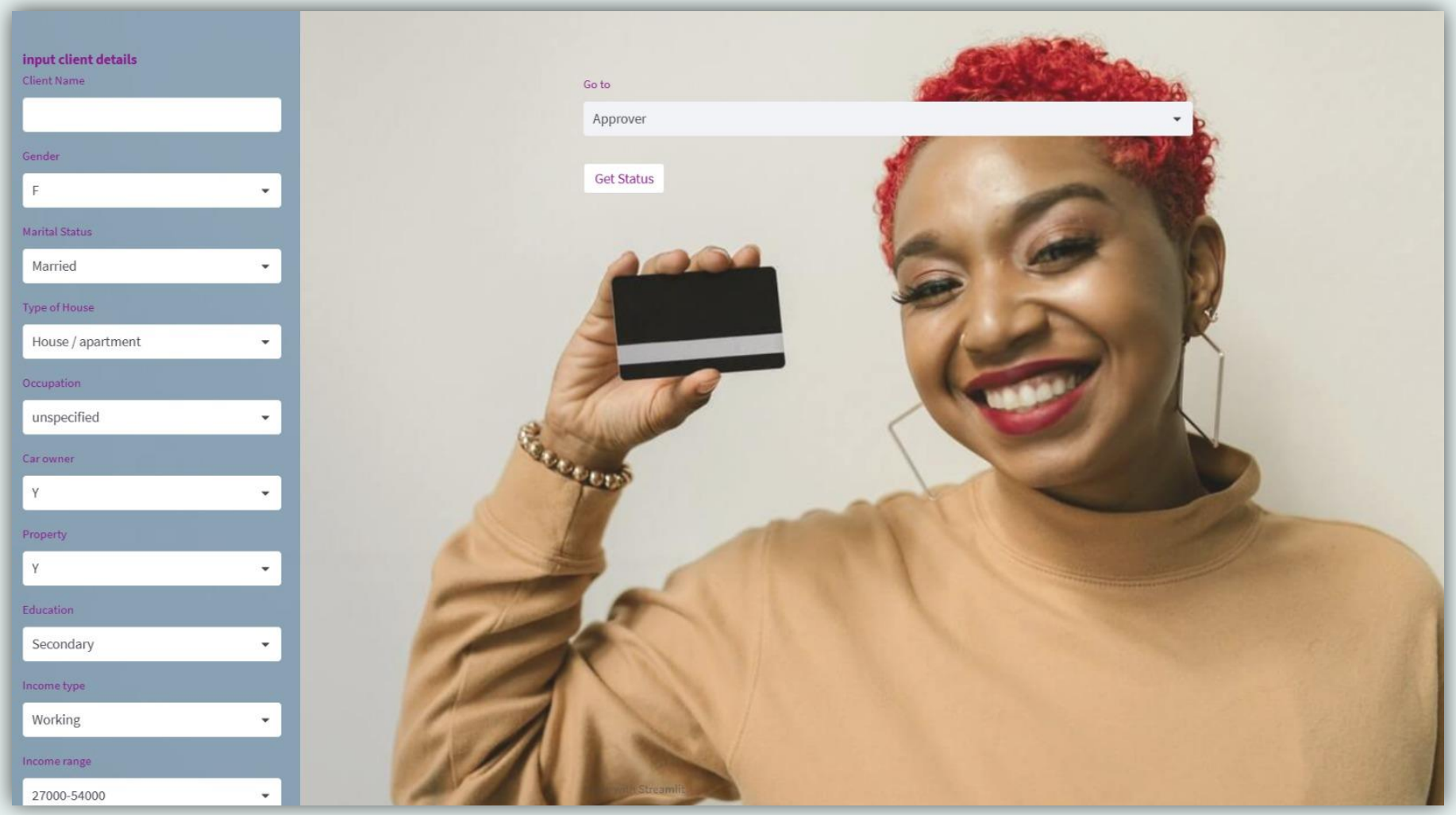
CONCLUSION

- The credit card approval system successfully addresses the limitations of traditional methods by leveraging machine learning algorithms and analyzing key applicant features, resulting in improved prediction accuracy.
- The system demonstrates high performance with a recall score of 0.85, indicating its ability to accurately classify approval or denial status and effectively manage risk.
- By incorporating advanced techniques and transparent decision-making, the system enables informed credit card approval decisions, leading to better risk management and increased efficiency in the credit scoring process.

RECOMMENDATIONS

- Conduct additional research or collect more data to better understand the demographic and social economic factors contributing to loan approval.
    - Explore the use of additional data sources, such as social media activity or online transaction history, to improve credit assessment and enhance predictive accuracy.
    - Establish a process for ongoing evaluation and refinement of the credit scoring model, including regular updates and retraining based on new data and feedback.

DEPLOYMENT


input client details
Client Name
Gender
F
Marital Status
Married
Type of House
House / apartment
Occupation
unspecified
Car owner
Y
Property
Y
Education
Secondary
Income type
Working
Income range
27000-54000
Go to
Approver
Get Status

THANK YOU!

ANY QUESTIONS?