CREDIT CARD DEFAULT PREDICTION

Presented By:

Teja kumar reddy Peddi - LD91652

INTRODUCTION

- Credit Defaulting: A significant challenge in today's financial landscape.
- Increase in Average American Credit Card Debt Up by 20% to 30% over the past decade.
- Importance of Customer Selection: Vital for credit card companies and the broader financial sector.
- Aim and Objective: Utilize data science and machine learning to predict credit card defaulting.
- Leveraging Demographic and Credit Factors: Analyzing various data points to make decisions.

DATASET

1. Sourced from the UCI Machine Learning Repository
2. Data from credit card clients in Taiwan from April 2005 to September 2005
3. Size: 2.86 MB
4. 30,000 Clients and 23 features
5. Contains information on default payments, demographic factors, credit data, and payment history

METHODOLOGY

Data Preparation

1. Data Gathering
2. Understanding data
3. Data Preprocessing or Cleaning

Data Analysis

1. Exploratory data analysis
2. Data Visualization
3. Extracting Key Insights

Data Modelling

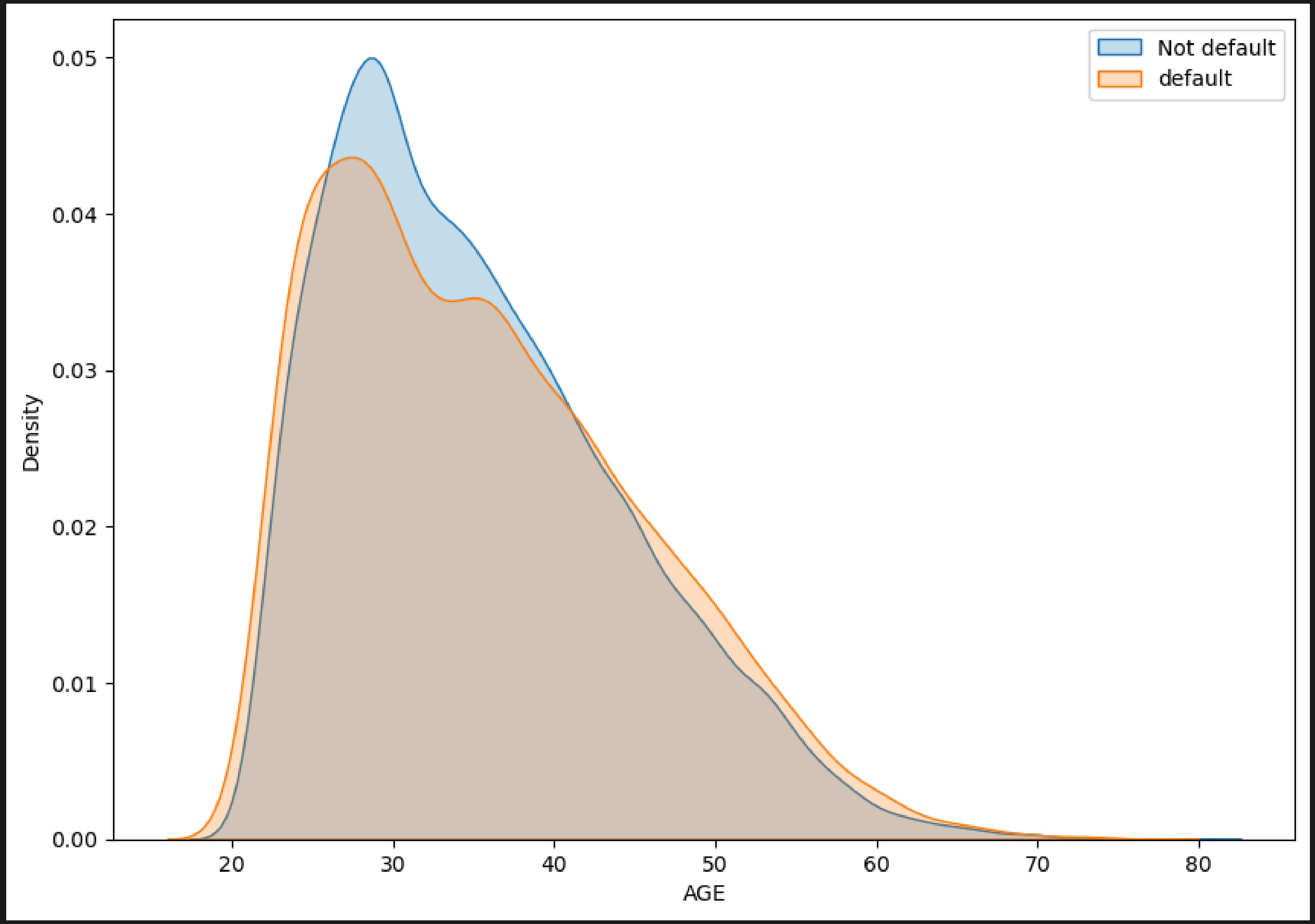
1. Model Fitting
2. Hyperparameter Tuning
3. Model Evaluation

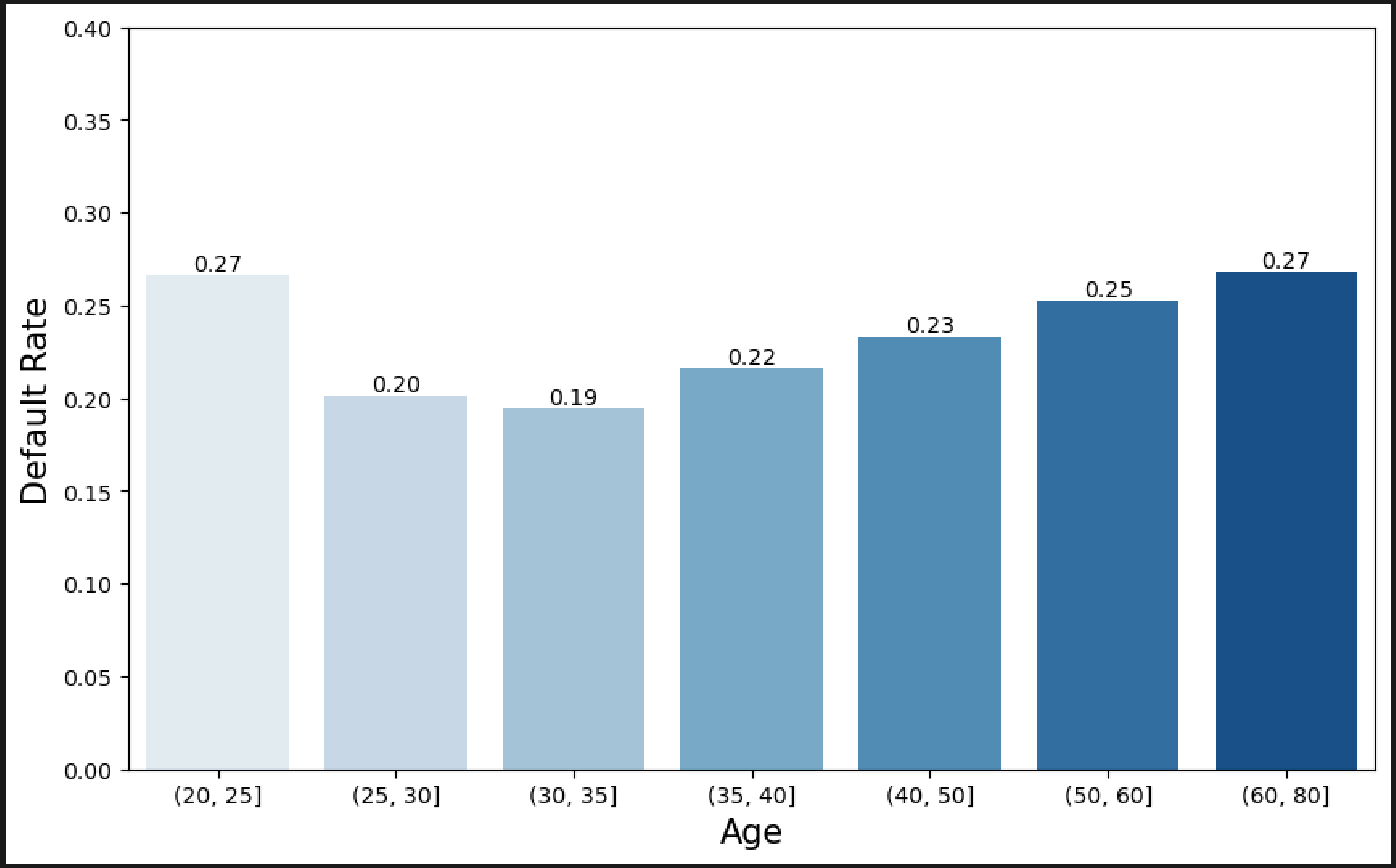
Deployment

1. Creating UI
2. Integrating model
3. Prediction

EXPLORATORY DATA ANALYSIS

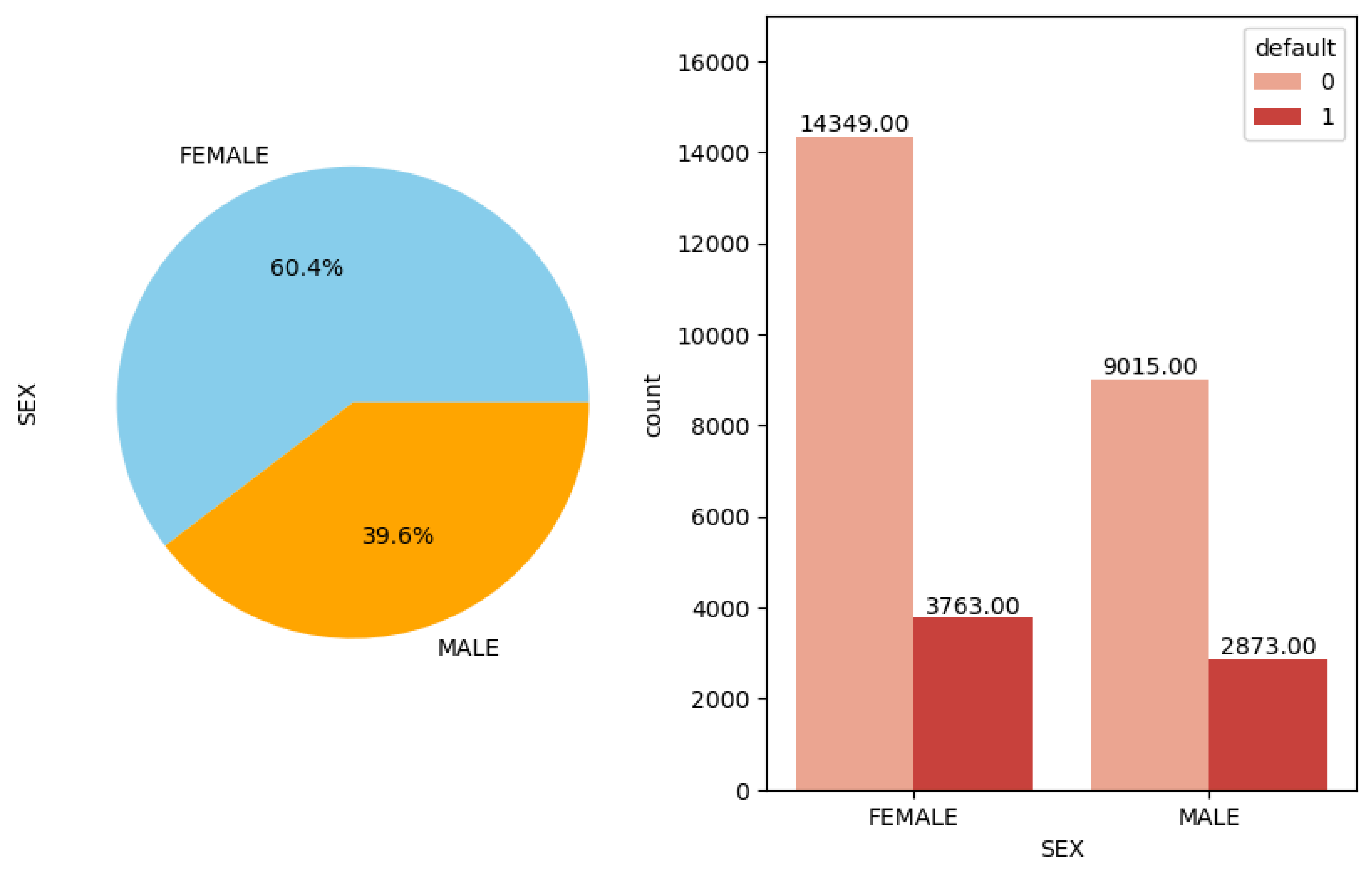
AGE

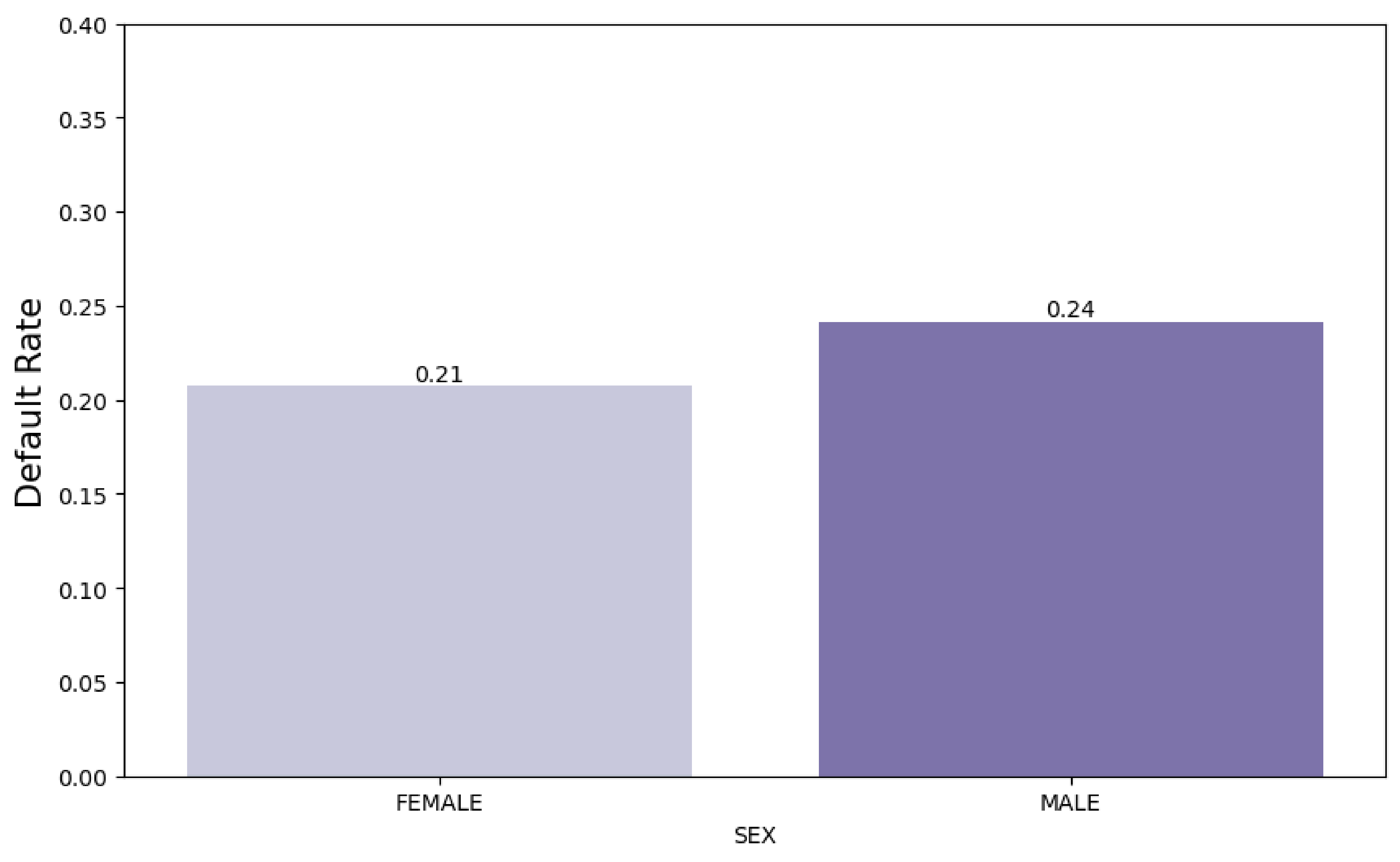




We can observe that the age group of 30-35 years old has lowest chances for defaulting, while the highest occur at the extremes(20-25 and 60+)

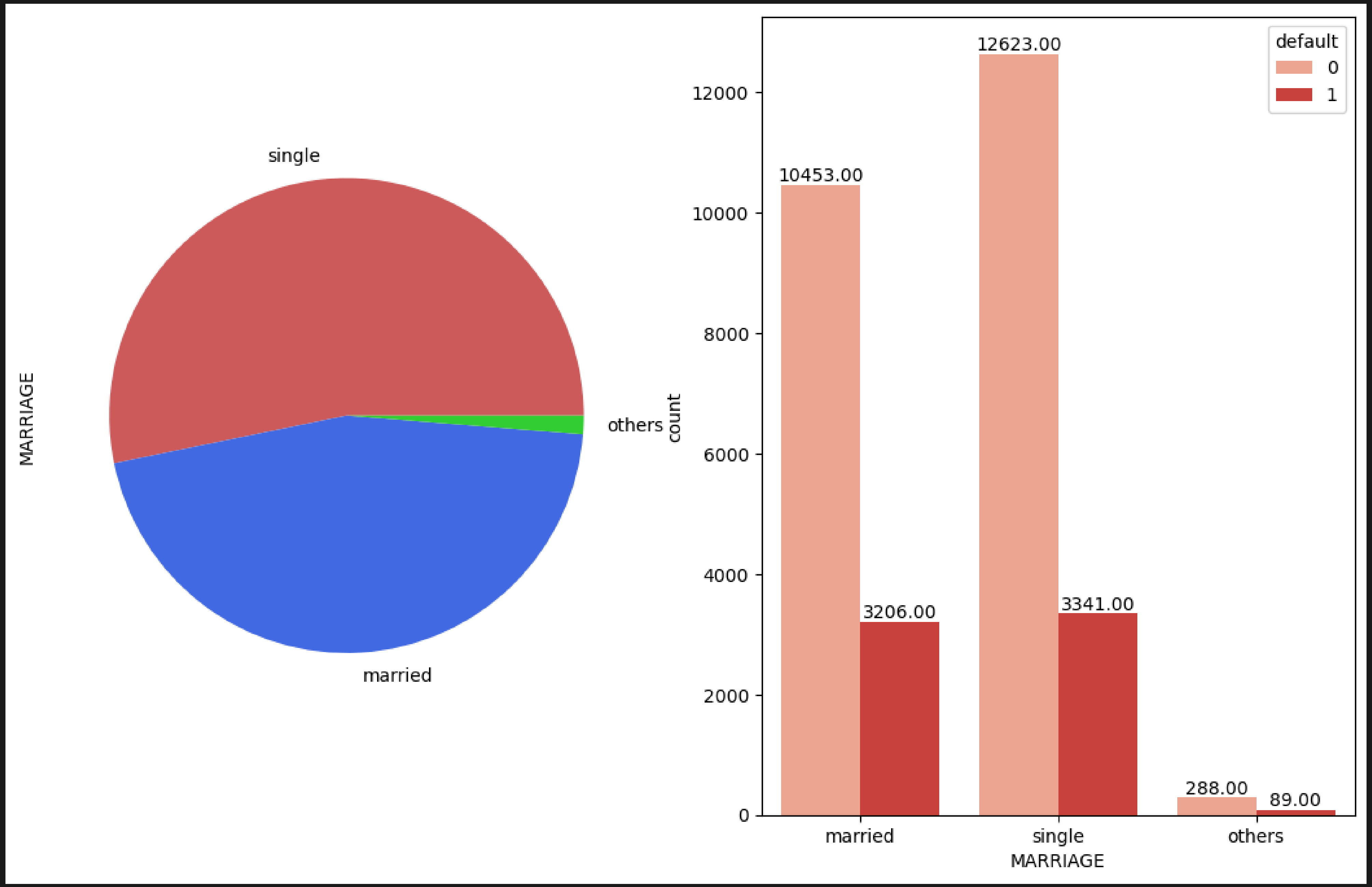
GENDER

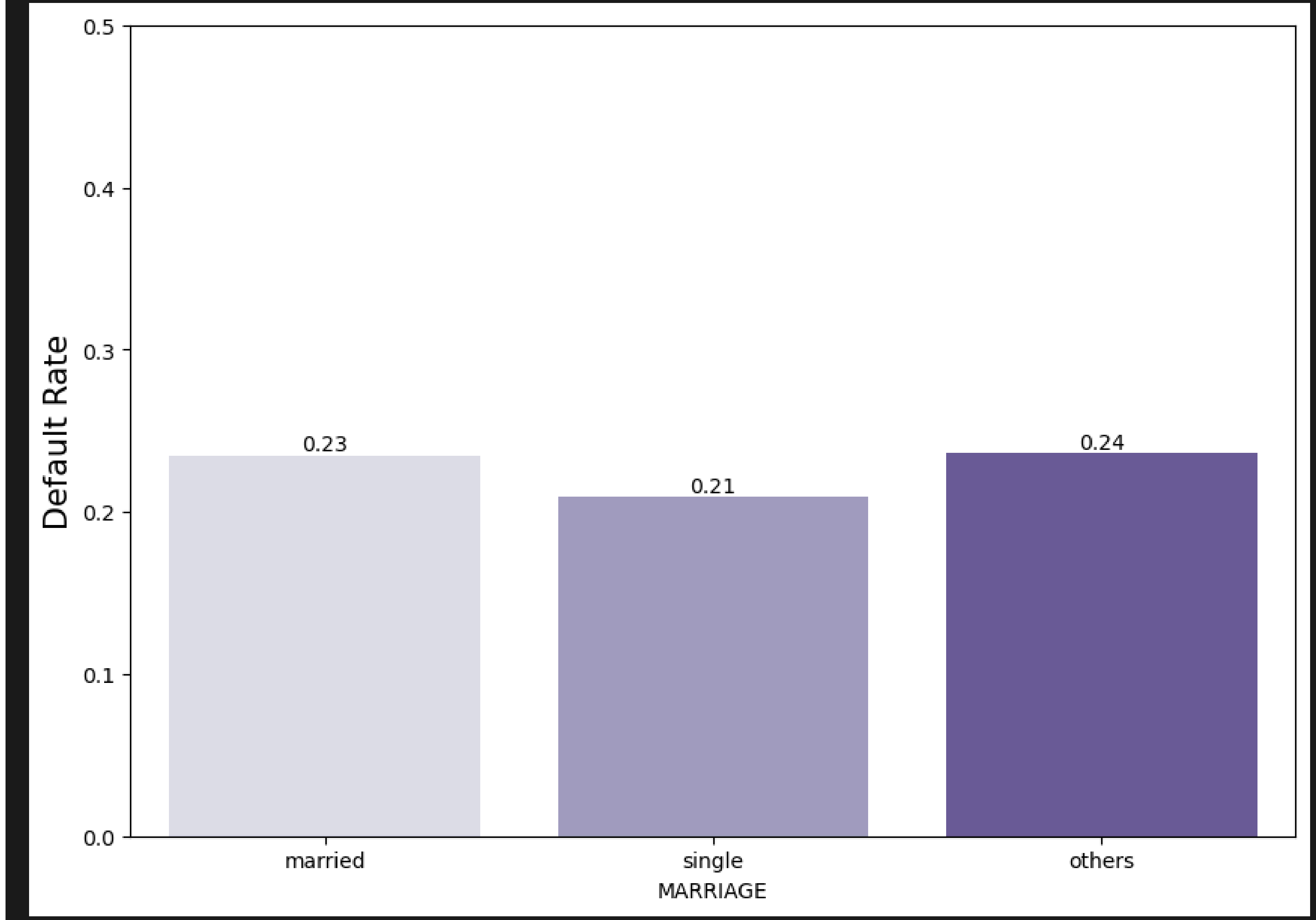




There are more women than men in the dataset and, men have a slightly higher chance of defaulting.

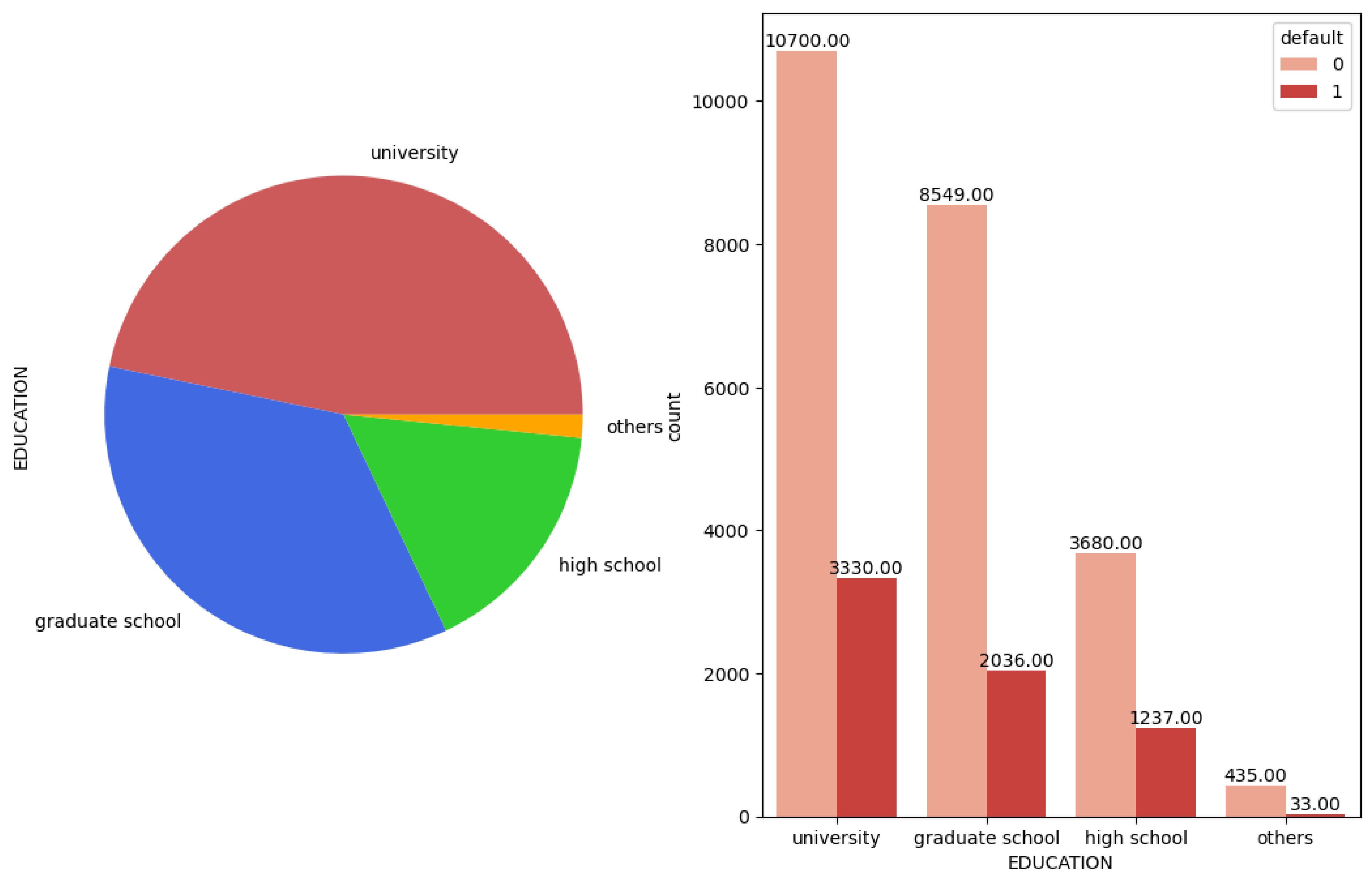
MARITIAL STATUS

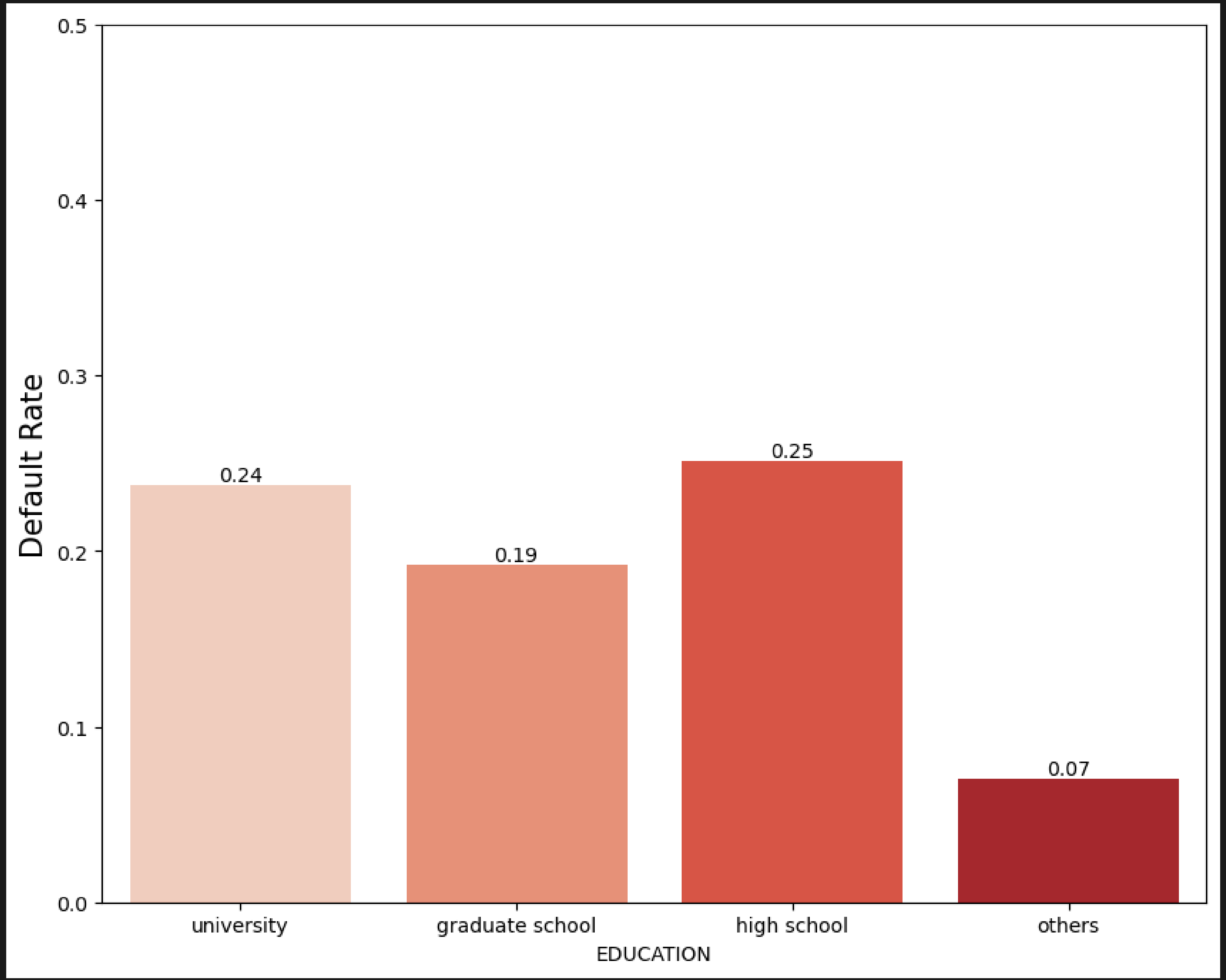




We can observe that Single people have lower chances of defaulting than married and other people.

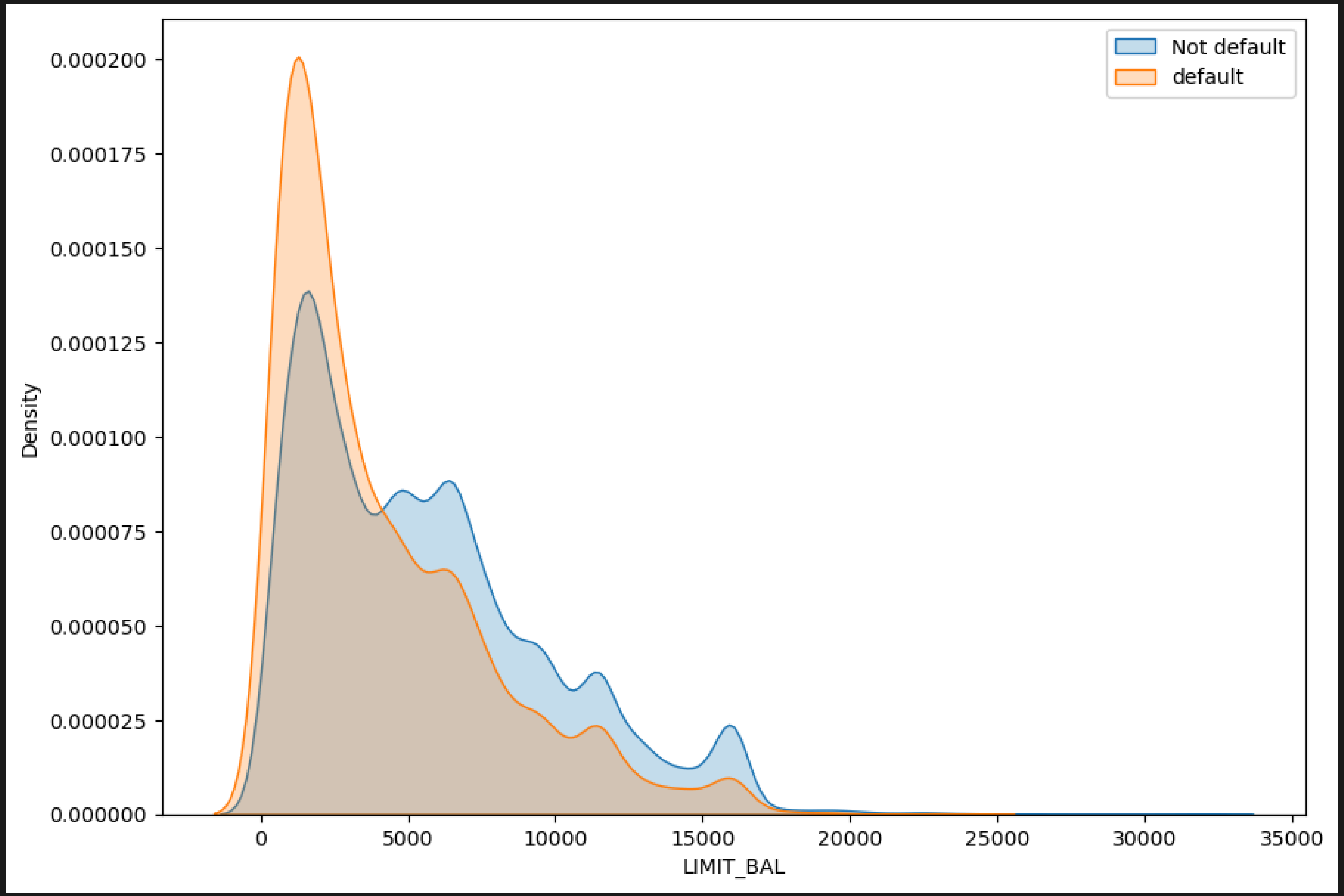
EDUCATION LEVEL

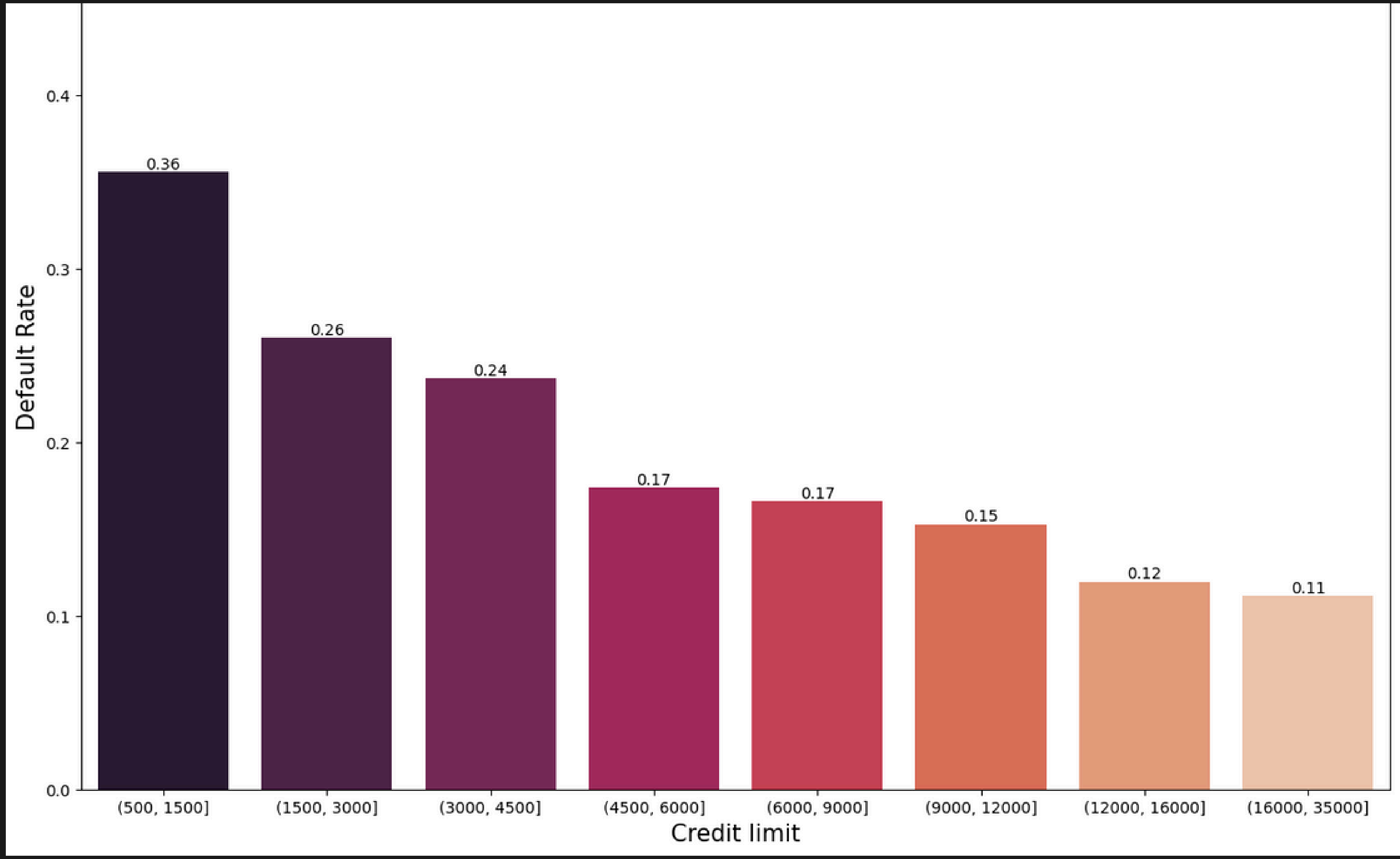




Considering the level of education, it seems that a higher education translates to a lower chance of default.

CREDIT LIMIT

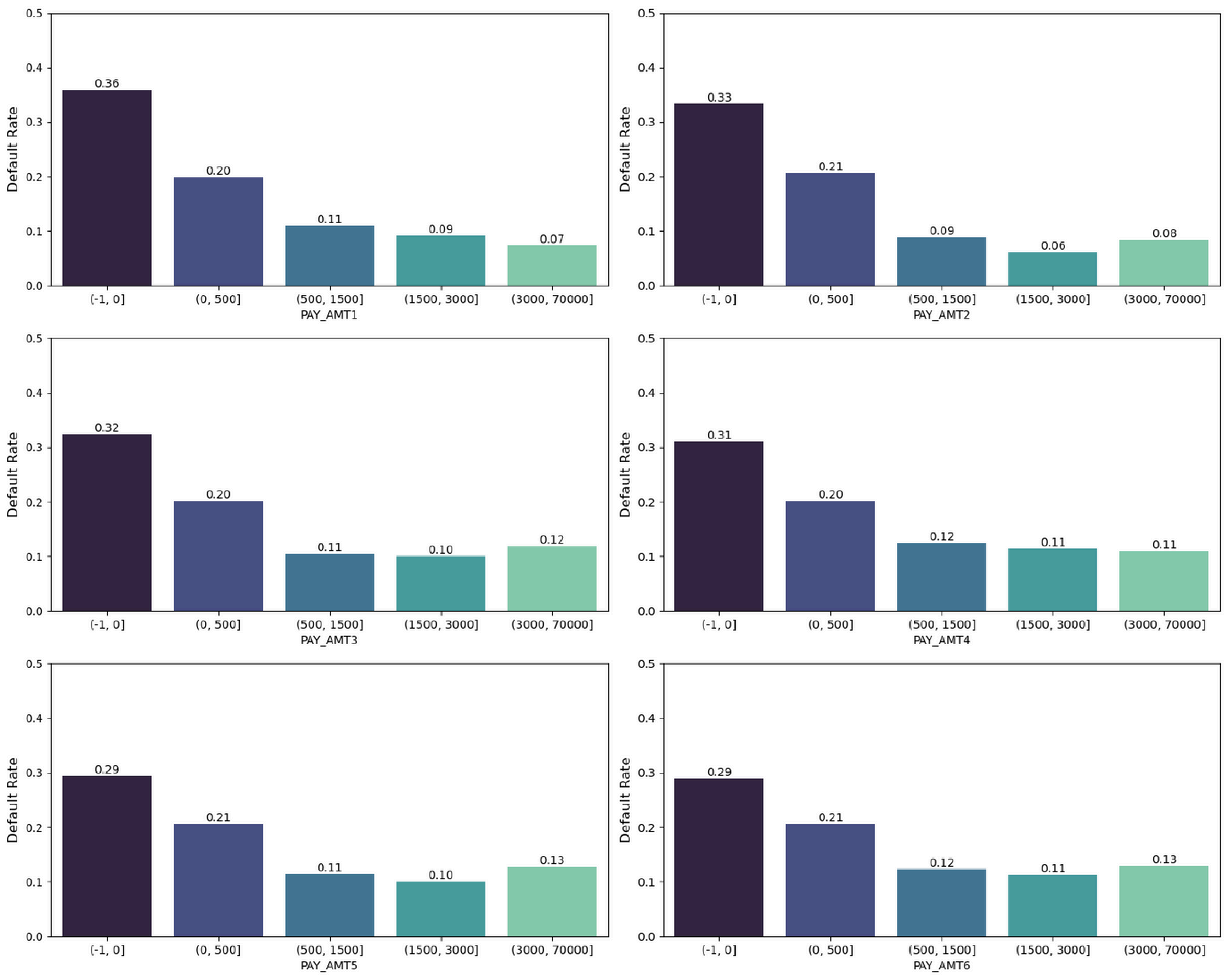




Most customers have a credit limit of 5K or less. Also it seems that a higher concentration defaulting occurs in the same range.

Over 30% of default can be observed with credit limit of 1500 USD or less

PREVIOUS PAYMENT



There is a higher default rate among those who paid nothing in previous months and a lower default rate among those who have done payment over 500 USD.

MODELLING

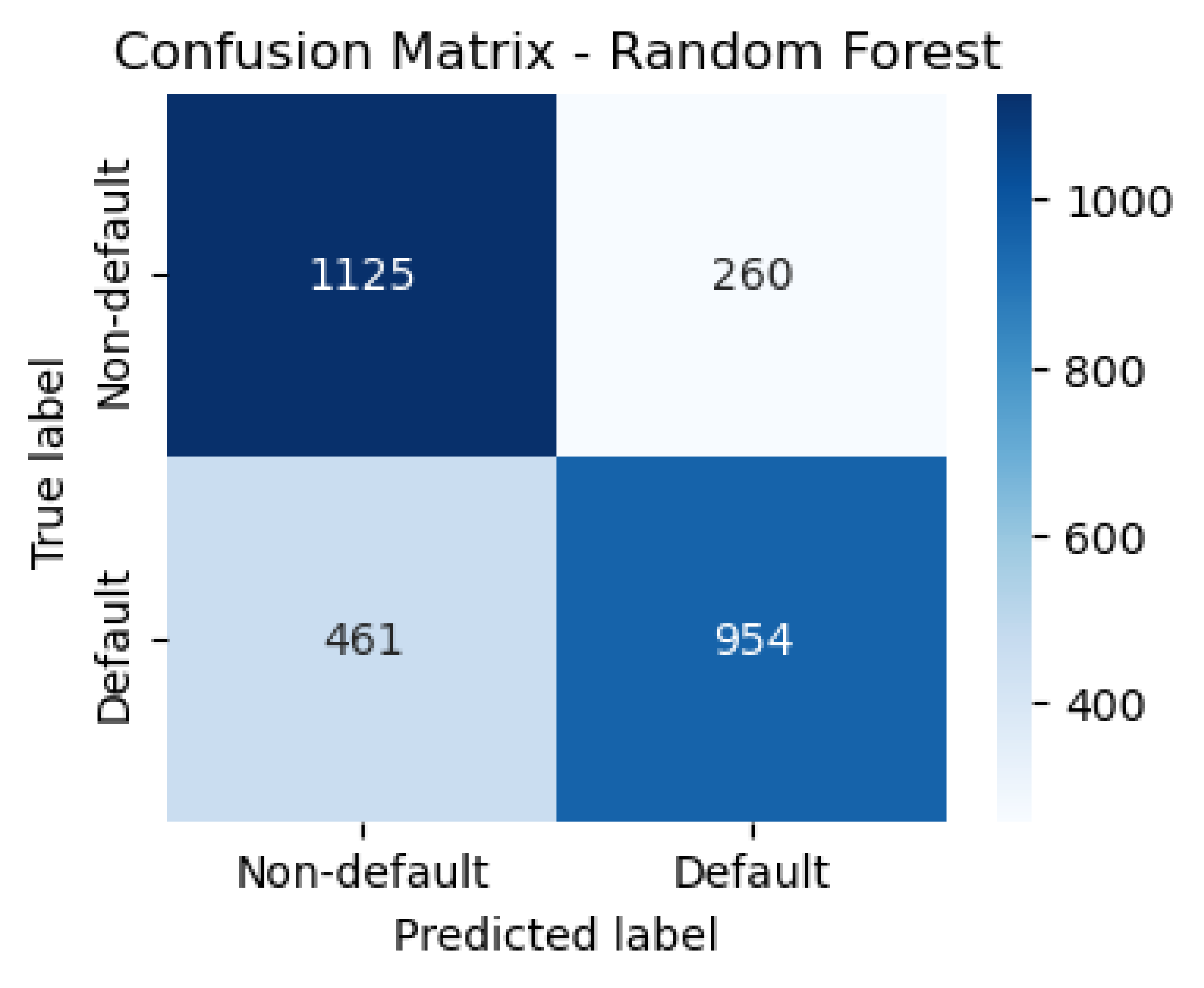
1. Resampled the data as the data was unbalanced
2. Train test split of 80-20
3. Implemented Logistic regression, Decision Trees, Random Forest Classifier and XGBoost
4. Achieved best results with Random Forest
5. Implemented Hyper parameter Tuning using GridSearchCV with 5-fold Cross validation
6. Saved the model

RESULTS

Best model: Random Forest Classifier

- Accuracy : 74.25%
- Precision : 0.75
- Recall : 0.74
- F1-Score : 0.74

1. **Accuracy**: Decent accuracy of 74.25% for an imbalanced problem
2. **Precision, Recall, and F1-Score**: The model performs better at identifying negatives (Class 0) than positives (Class 1).
3. **Confusion Matrix Insights**: There is a decent balance of values for general predictability but there is space for improvement for false negatives which can be achieved using better data



FUTURE SCOPE

1. **Incorporating More Data**: A balanced data and an adequate amount of samples of both the classes can significantly improve the model's accuracy and predictive power.
2. **Real-Time Prediction Systems**: Developing real-time prediction capabilities will allow for instant risk assessments, enabling financial institutions to make timely and informed decisions.
3. **Continuous Monitoring and Recalibration**: Regularly updating and recalibrating the model will ensure it remains effective and accurate in response to changing economic conditions and consumer behaviors.

CONCLUSION

- Overall, the trained machine learning model offers a robust tool for financial institutions to manage credit risk effectively.
- By identifying high-risk customers through key features like payment history and credit utilization, the model enables proactive risk management and more informed lending decisions.
- This predictive capability can lead to reduced default rates and improved financial stability.
- Future enhancements and real-time implementations promise even greater accuracy and adaptability to changing economic conditions

THANKYOU