

**CREDIT EDA ASSIGNMENT**

# **EXPLORATORY DATA ANALYSIS CASE STUDY**

---

# PROBLEM STATEMENT:

Consumer finance companies face challenges in loan approval decisions due to insufficient credit histories of applicants. This leads to two key risks:

- **Loss of Business:** Rejecting applicants who could repay the loan results in lost opportunities.
- **Financial Loss:** Approving loans for applicants likely to default can cause financial damage.

To address these issues, the company needs to use Exploratory Data Analysis (EDA) to analyse loan application data, identify patterns indicative of loan default, and improve decision-making processes.

---

---

# ASSIGNMENT OBJECTIVE

The objective of this assignment is to leverage EDA to:

- **Identify Patterns:** Uncover key attributes and patterns that predict loan default.
  - **Assess Risk:** Determine driving factors behind loan defaults to inform decision-making on loan approvals, amounts, and interest rates.
  - **Enhance Decisions:** Reduce financial risks and avoid rejecting applicants capable of repaying loans by using insights from the analysis.
-

---

# STEPS TO FOLLOW:

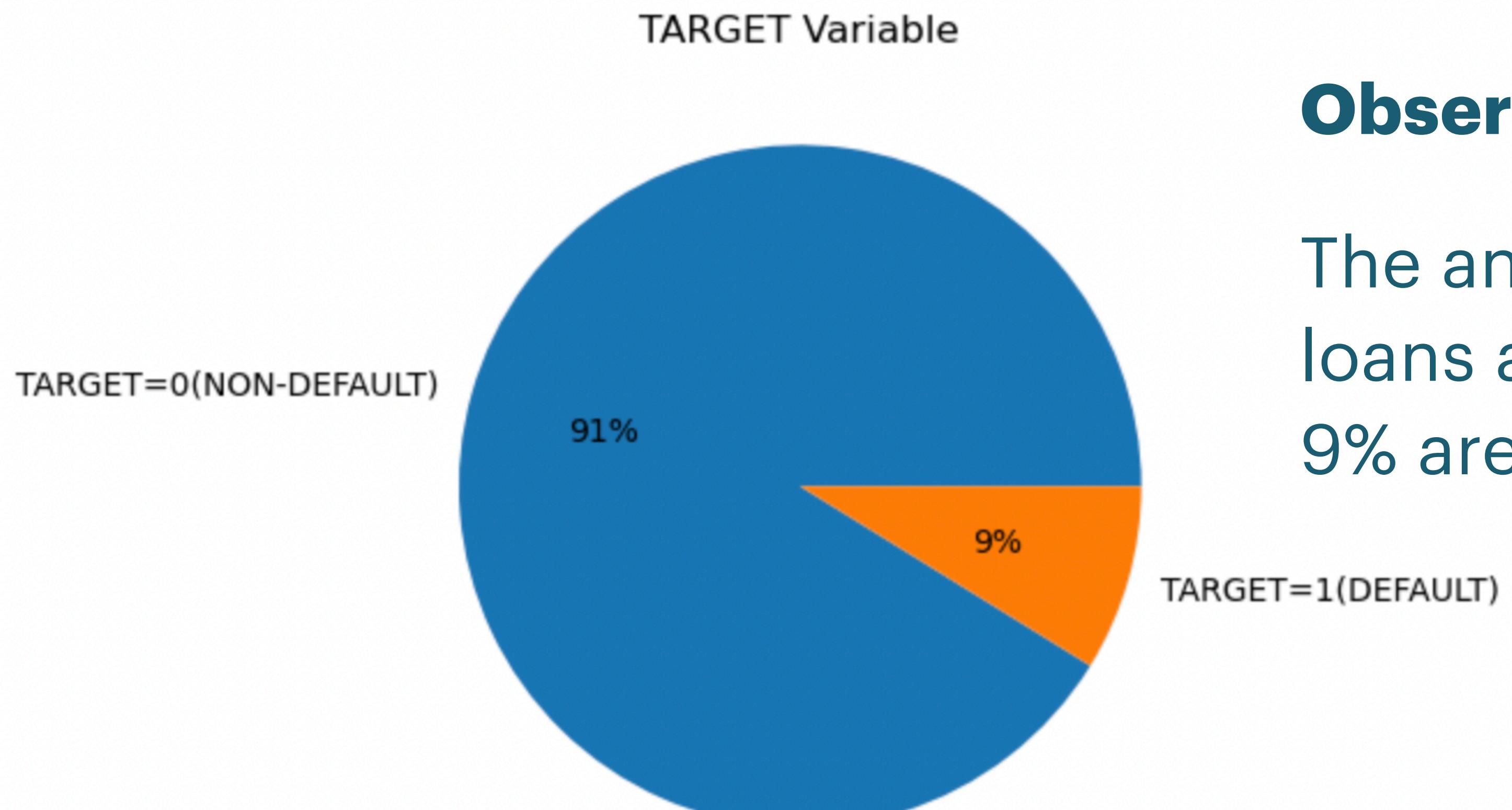
- **Understand the Problem and Data Structure:** Familiarise yourself with the problem and how the data is organised.
  - **Verify Data Quality:** Assess the accuracy and completeness of the data.
  - **Address Missing Data:** Decide whether to remove missing data or fill in gaps with appropriate values.
  - **Detect Outliers:** Identify and handle any outliers in numerical data.
  - **Bin Data for Analysis:** Group data into bins for more manageable analysis.
  - **Identify Target Attribute and Check for Imbalance:** Determine the target variable and assess whether the data is imbalanced.
  - **Segment Data by Target Attribute:** Divide the data into two groups based on the target attribute (e.g., defaulters vs. non-defaulters).
  - **Conduct Univariate and Bivariate Analysis:** Analyze individual variables and their relationships, and determine correlations.
  - **Integrate Data:** Combine data from previous and current applications for comprehensive analysis.
  - **Assess Risk and Provide Recommendations:** Identify potential risks and suggest improvements.
-

---

# **DATA ANALYSIS OVERVIEW**

---

# ANALYSING DATA IMBALANCE



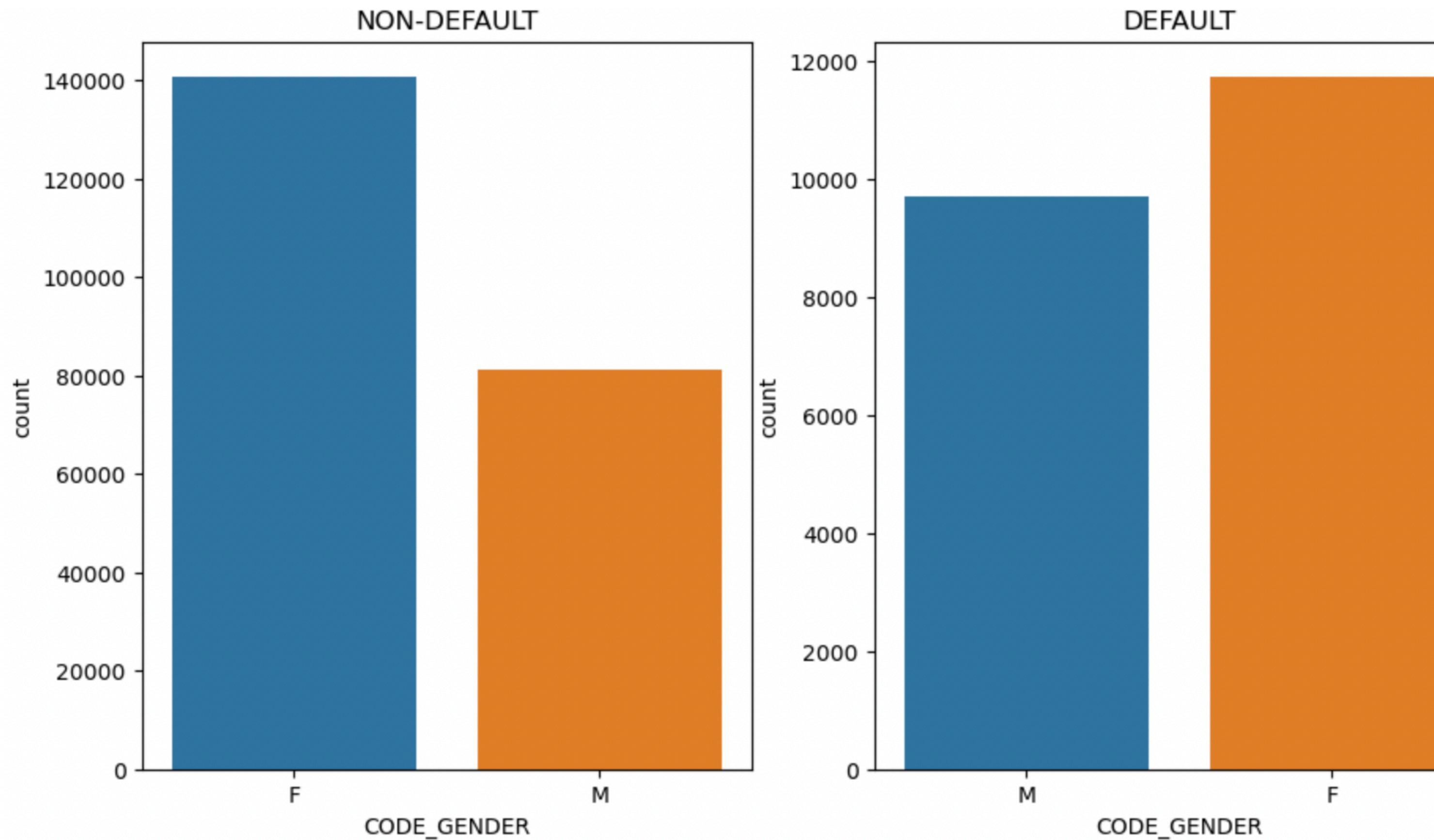
## Observation:

The analysis reveals that 91% of the loans are non-defaulters, while only 9% are defaulters.

# UNIVARIATE ANALYSIS

w.r.t TARGET=0 & TARGET=1

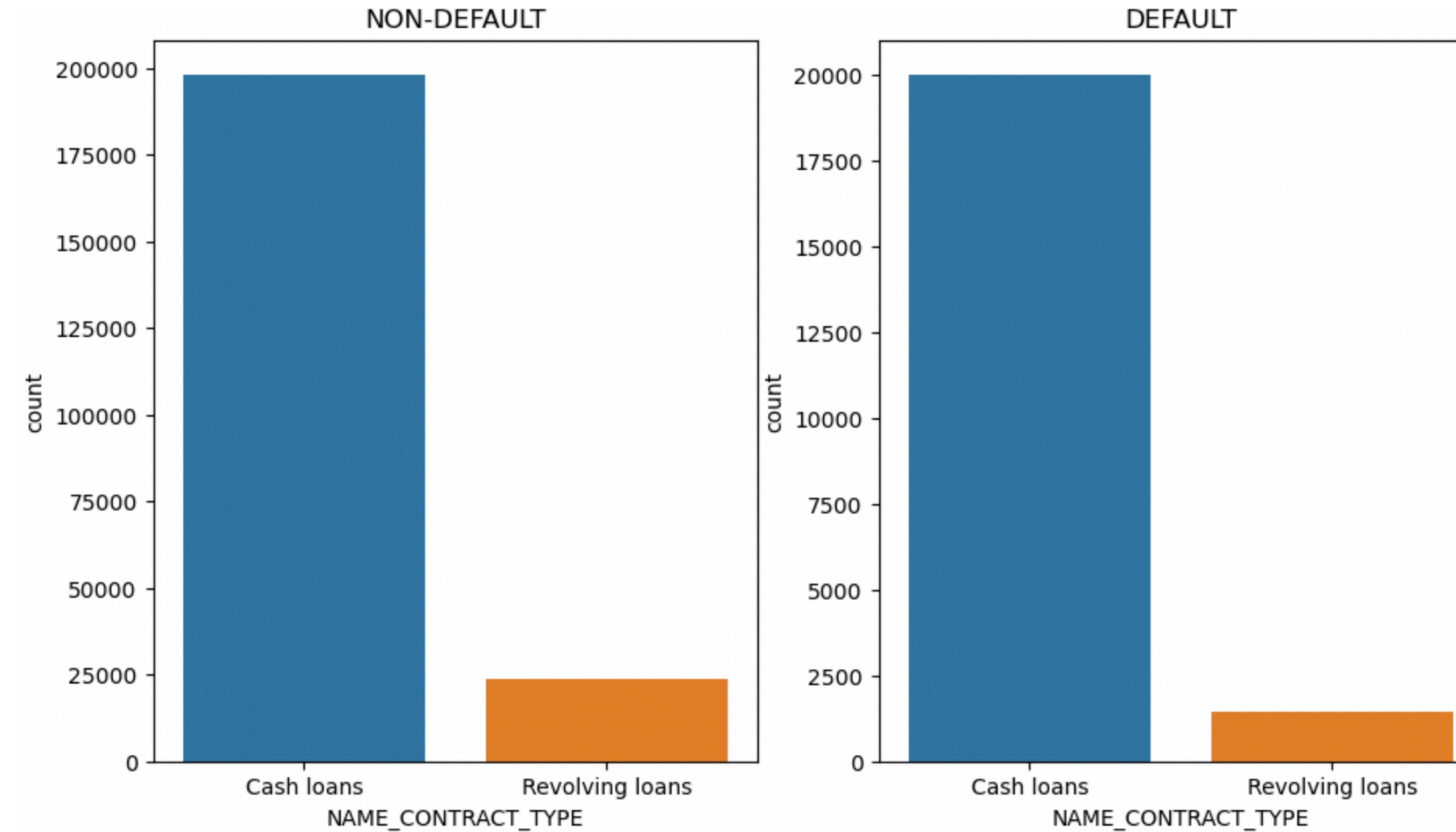
## 1.1 Analysis of "CODE\_GENDER" column



### Observation:

There is a higher number of female loan applicants compared to males, resulting in a higher number of female defaulters. Additionally, the default rate among male applicants is higher compared to non-defaulters.

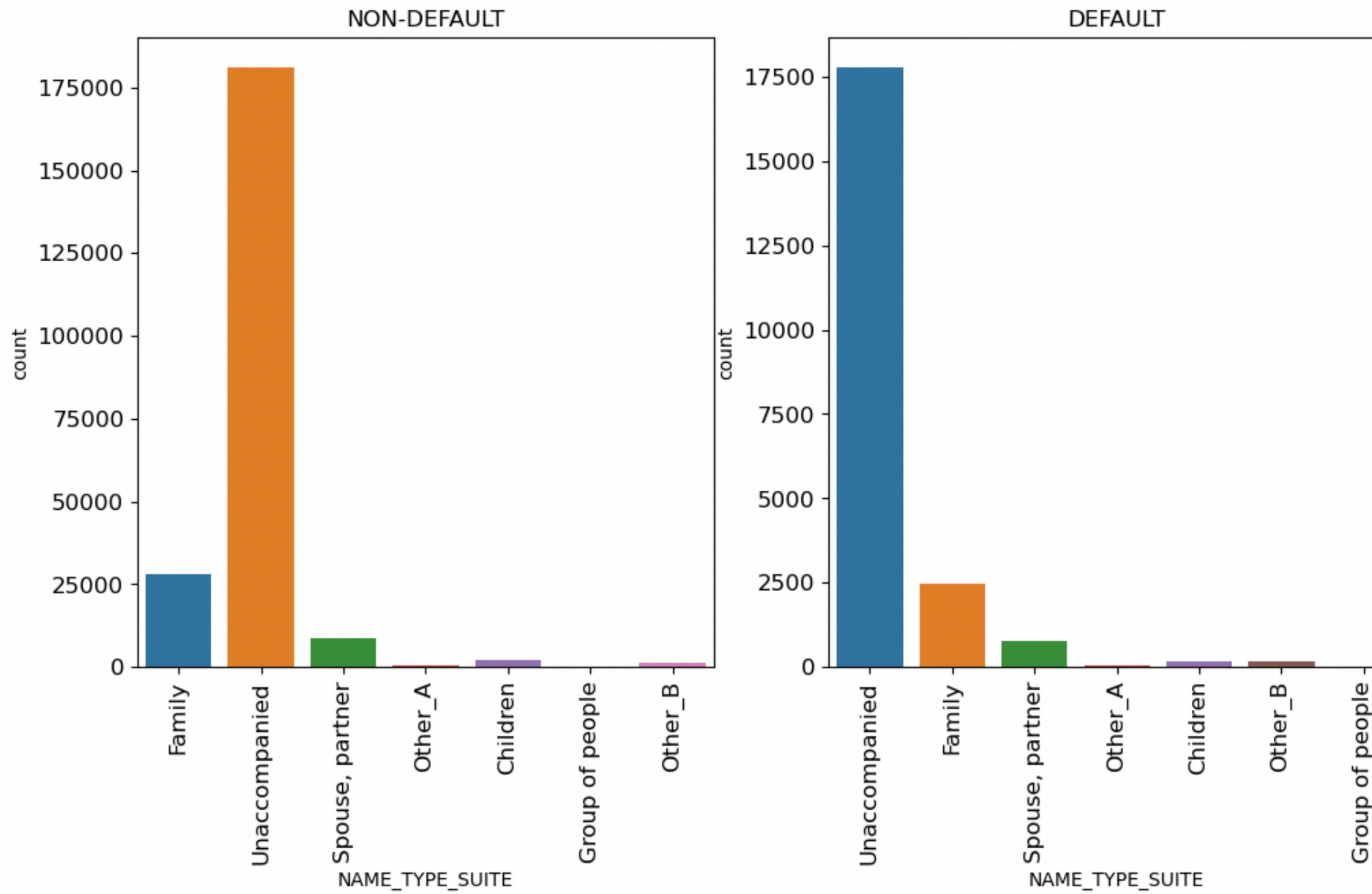
## 1.2 Analysis of "NAME\_CONTRACT\_TYPE" column



### Observation:

The data indicates that revolving loans are less common among defaulted loans. This suggests that revolving loans might be safer, likely due to their flexible repayment and borrowing features.

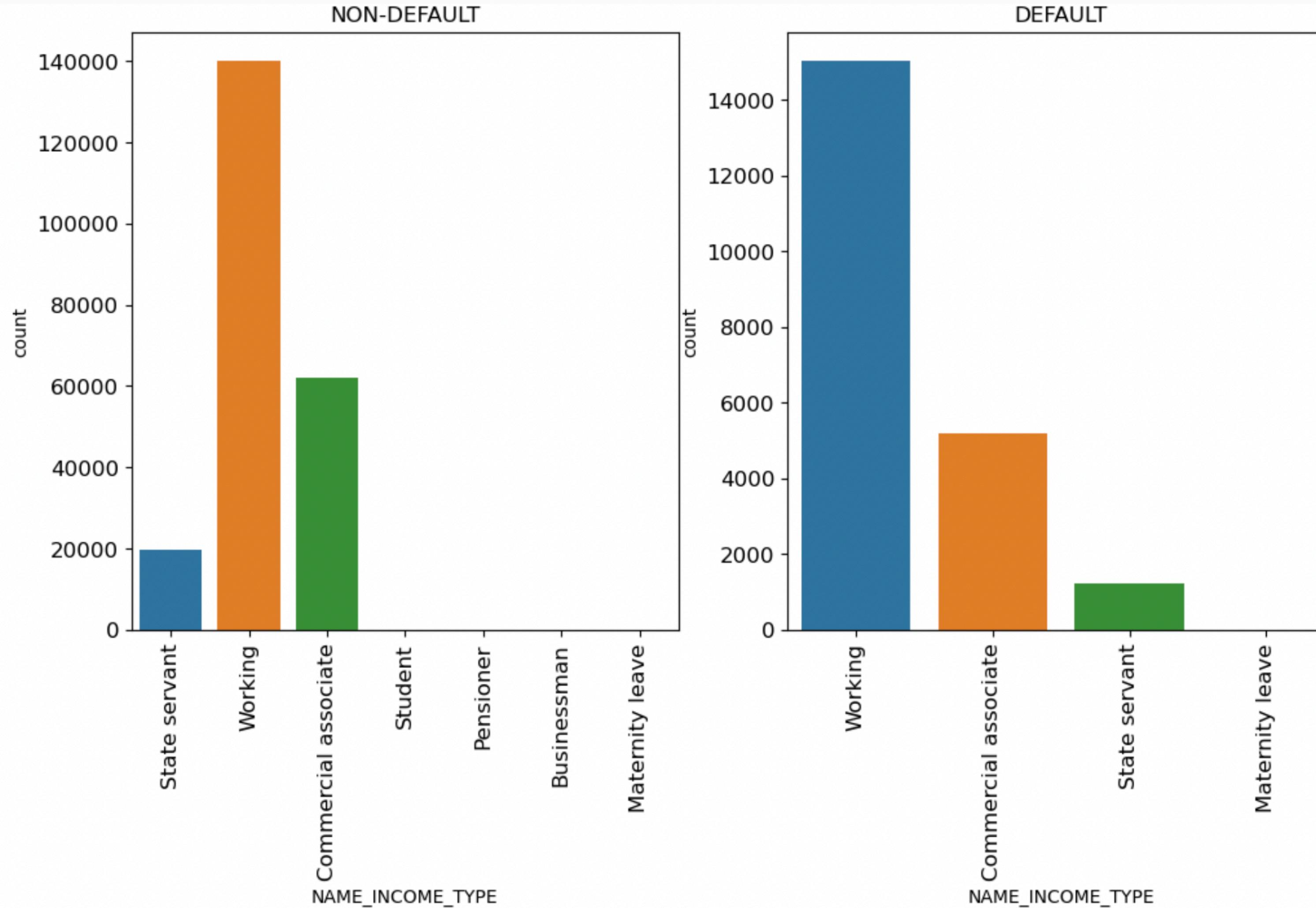
### 1.3 Analysis of "NAME\_TYPE\_SUITE" column



#### Observation:

The data indicates that revolving loans are less common among defaulted loans. This suggests that revolving loans might be safer, likely due to their flexible repayment and borrowing features.

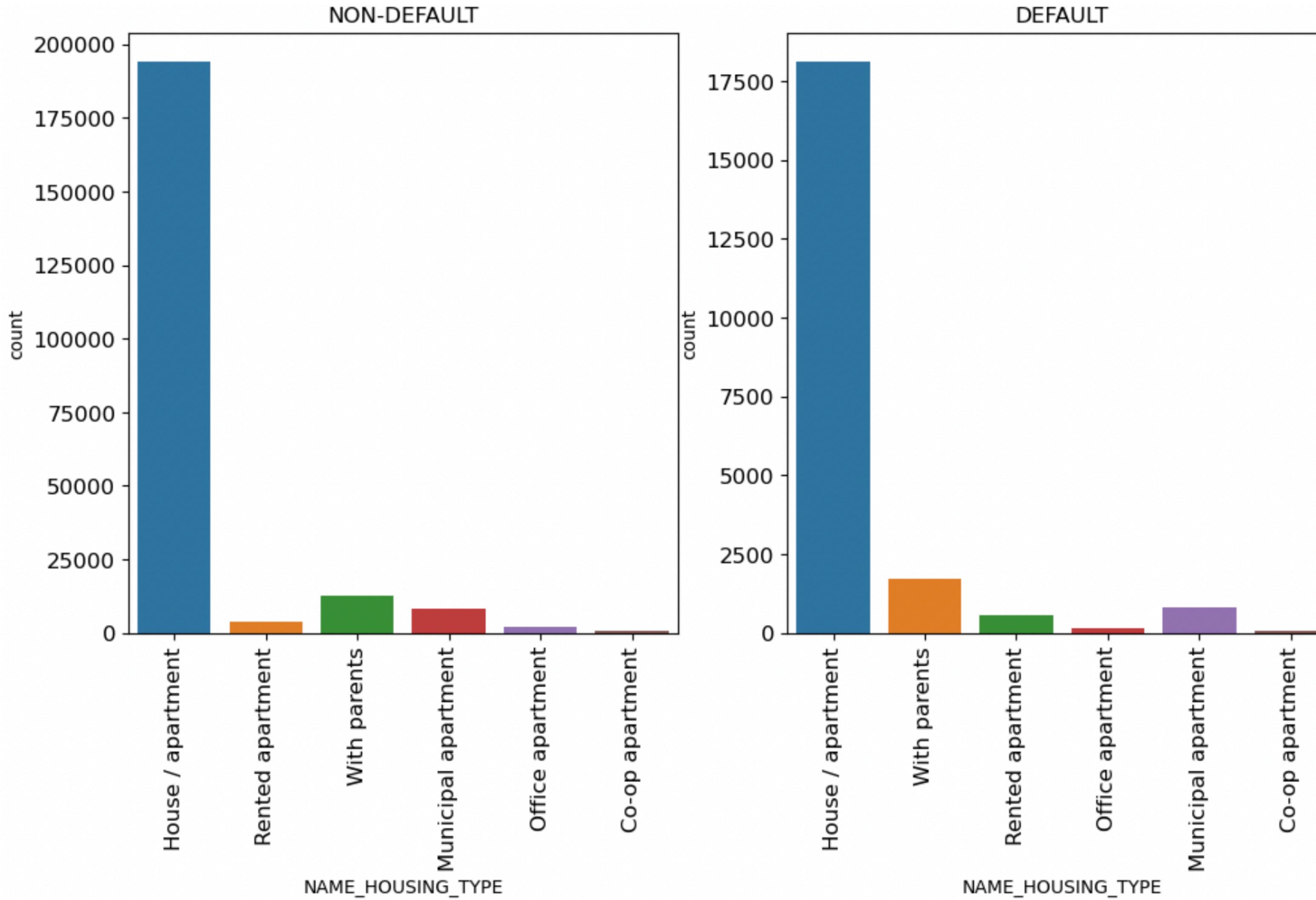
## 1.4 Analysis of “NAME\_INCOME\_TYPE” column



### Observation:

The majority of loan defaults are seen among the working population, compared to other income groups.

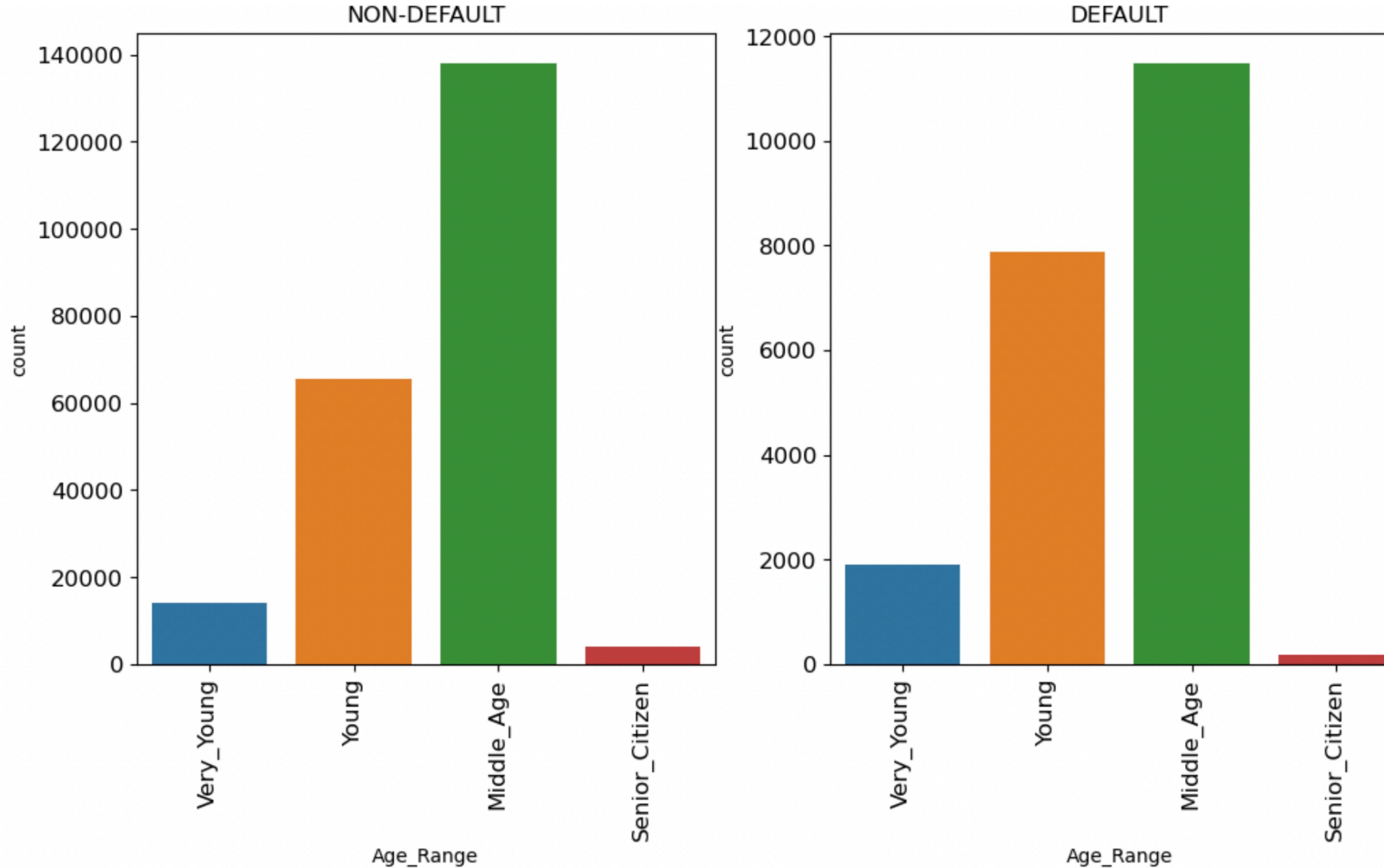
## 1.5 Analysis of “NAME\_HOUSING\_TYPE” column



### Observation:

Individuals who own a house or apartment are more likely to apply for loans. However, those living with their parents are more prone to default compared to other housing situations.

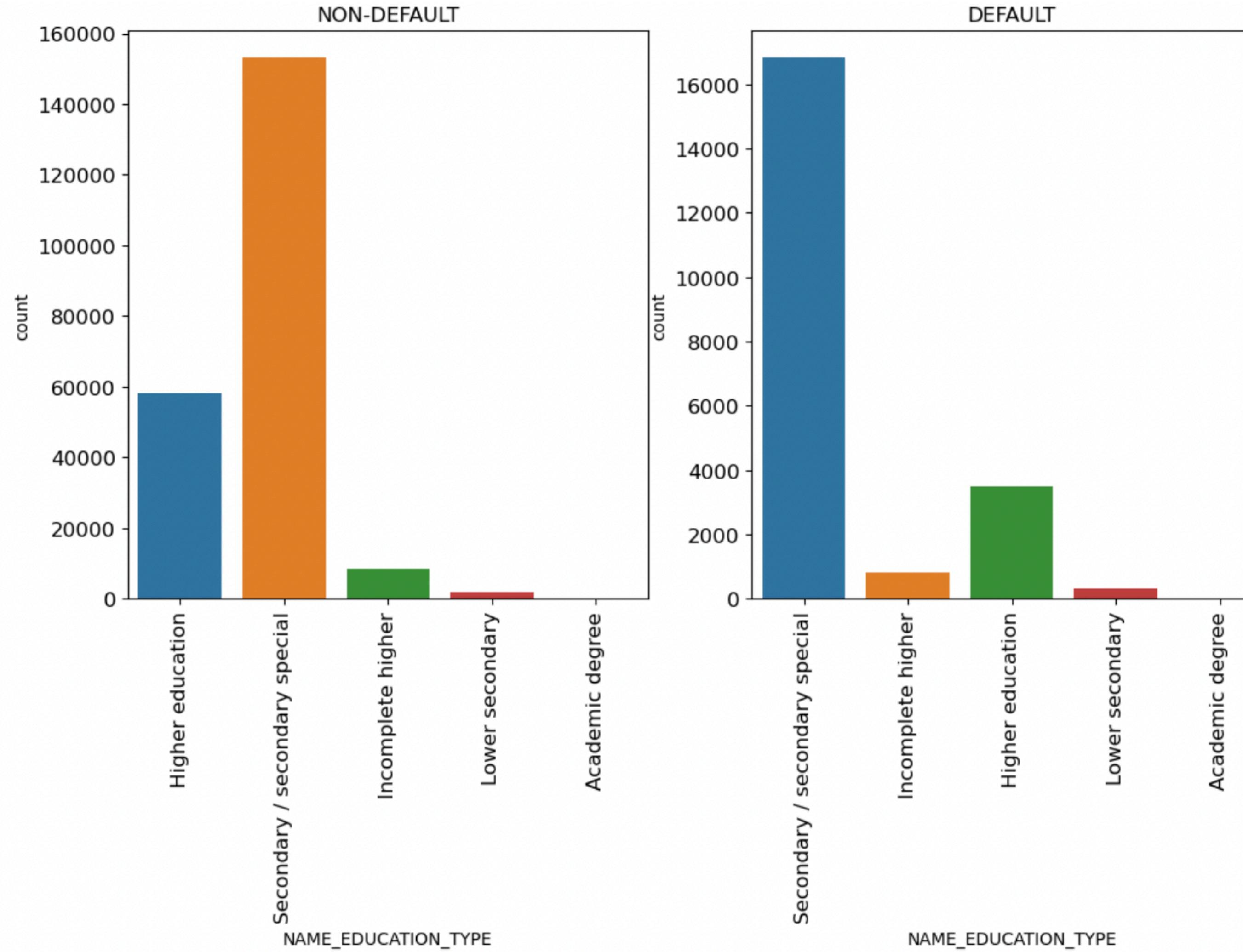
## 1.6 Analysis of “Age\_Range” column



### Observation:

The middle-aged group is the most frequent loan applicant and also represents the highest risk of default compared to other age groups.

## 1.7 Analysis of “NAME\_EDUCATION\_TYPE” column

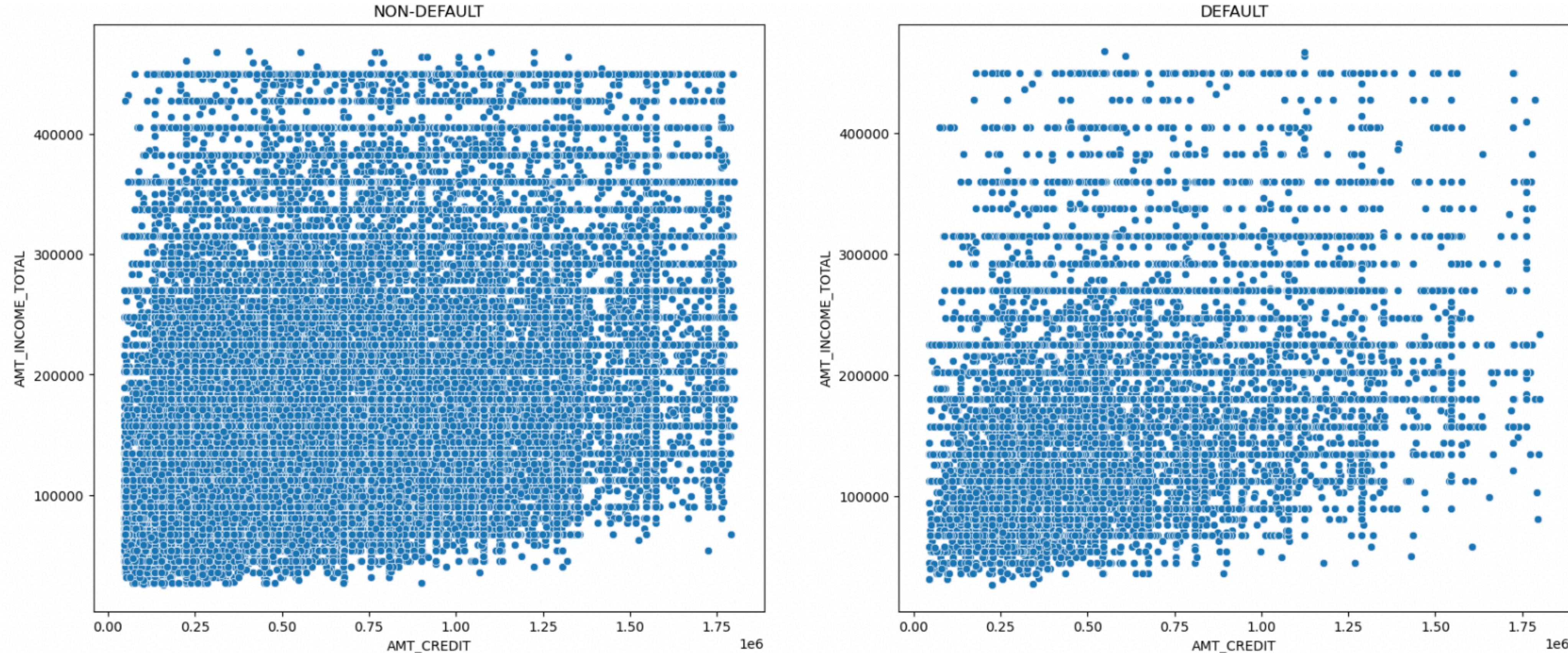


### Observation:

People with a secondary education level are the most frequent loan applicants and also have the highest rate of defaults compared to other educational levels.

# BIVARIATE ANALYSIS

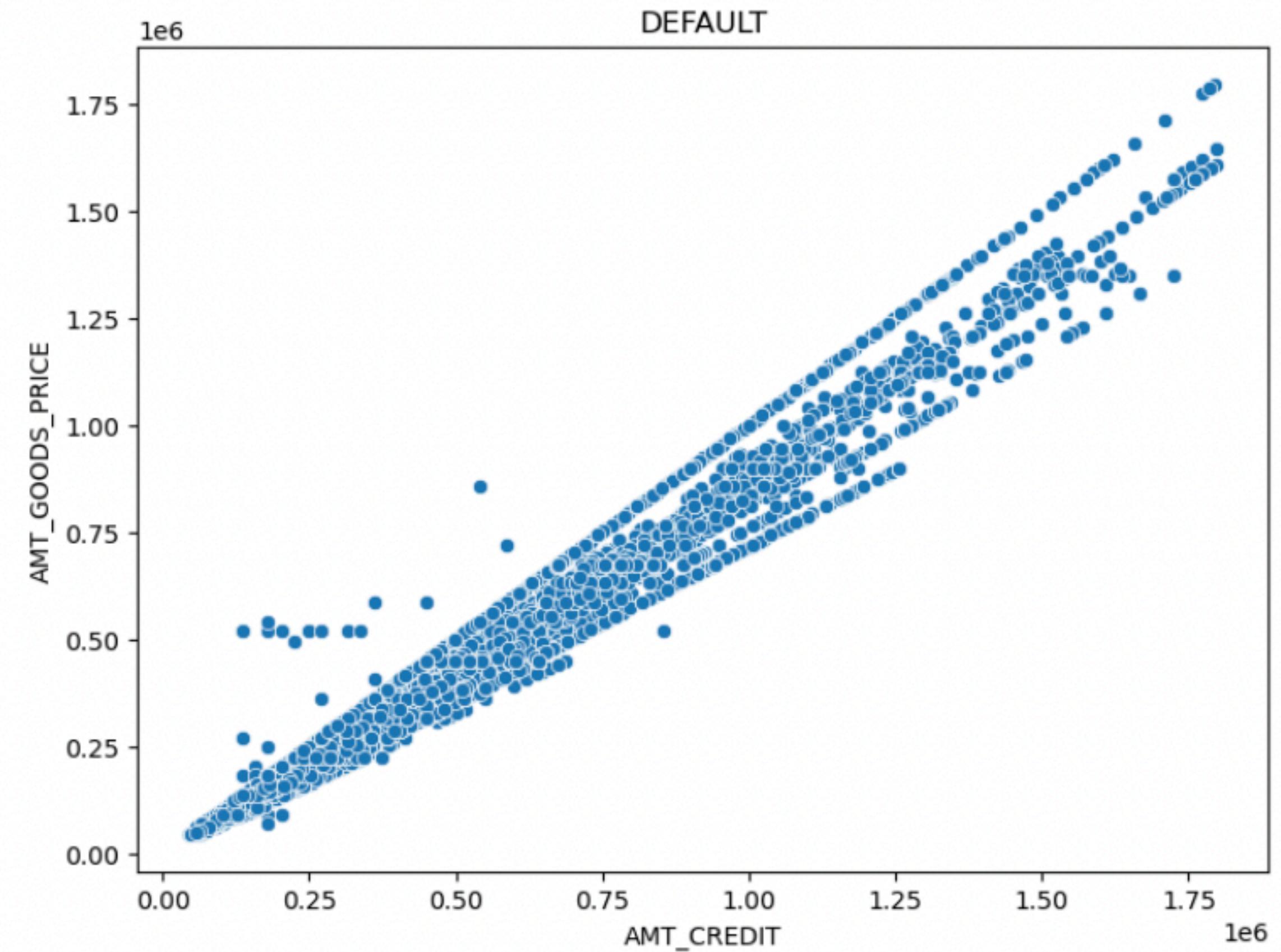
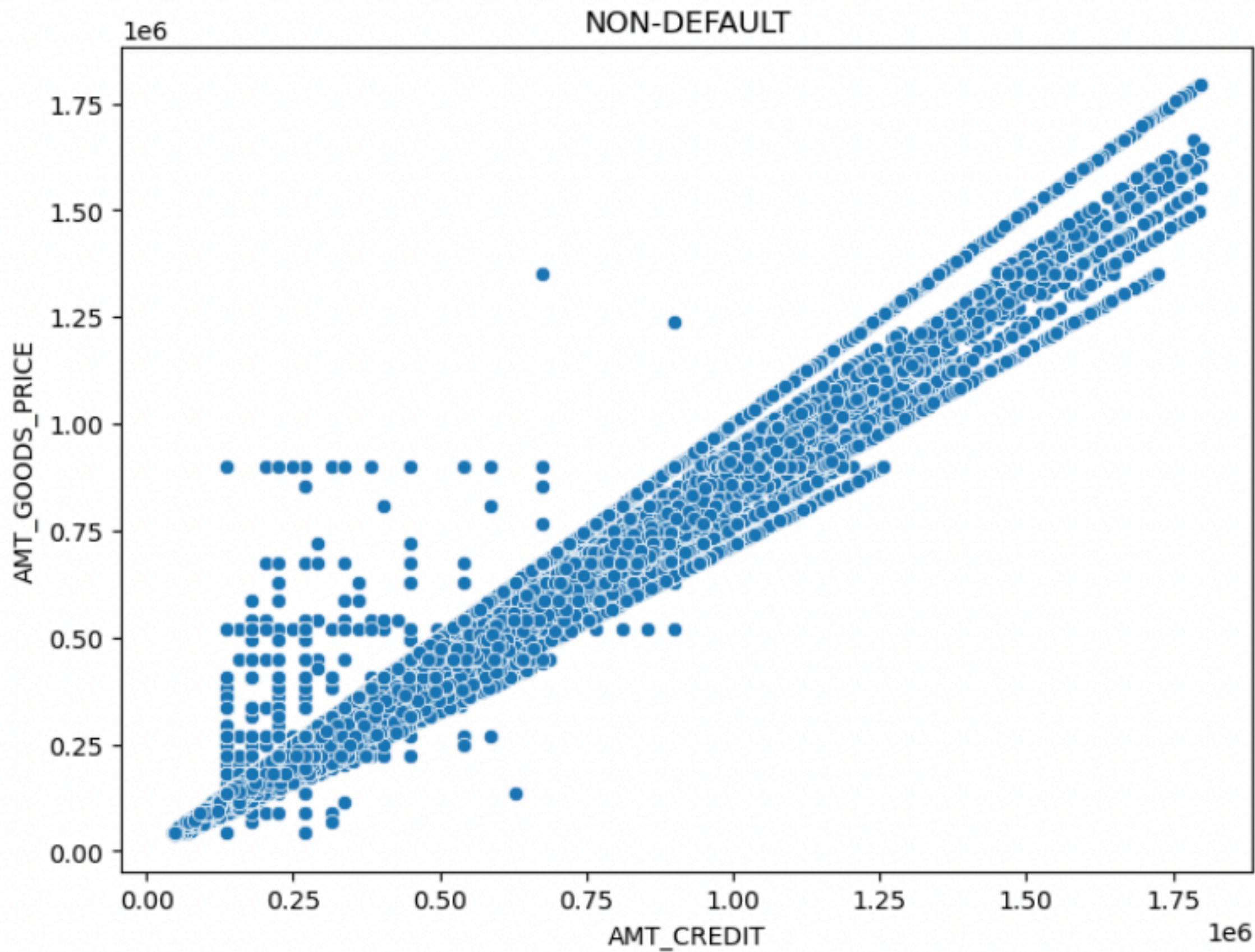
## 1.1 Analysis of "AMT\_CREDIT" & "AMT\_INCOME\_TOTAL"



### Observation:

There is a lower density of defaults when the income exceeds 300k or when the credit amount is below 200k.

### 1.2 Analysis of “AMT\_CREDIT” & “AMT\_GOODS\_PRICE”



#### Observation:

Defaults are less frequent when the price of goods is up to 500k and the credit amount is also below 500k.

# COMBINED DATA ANALYSIS

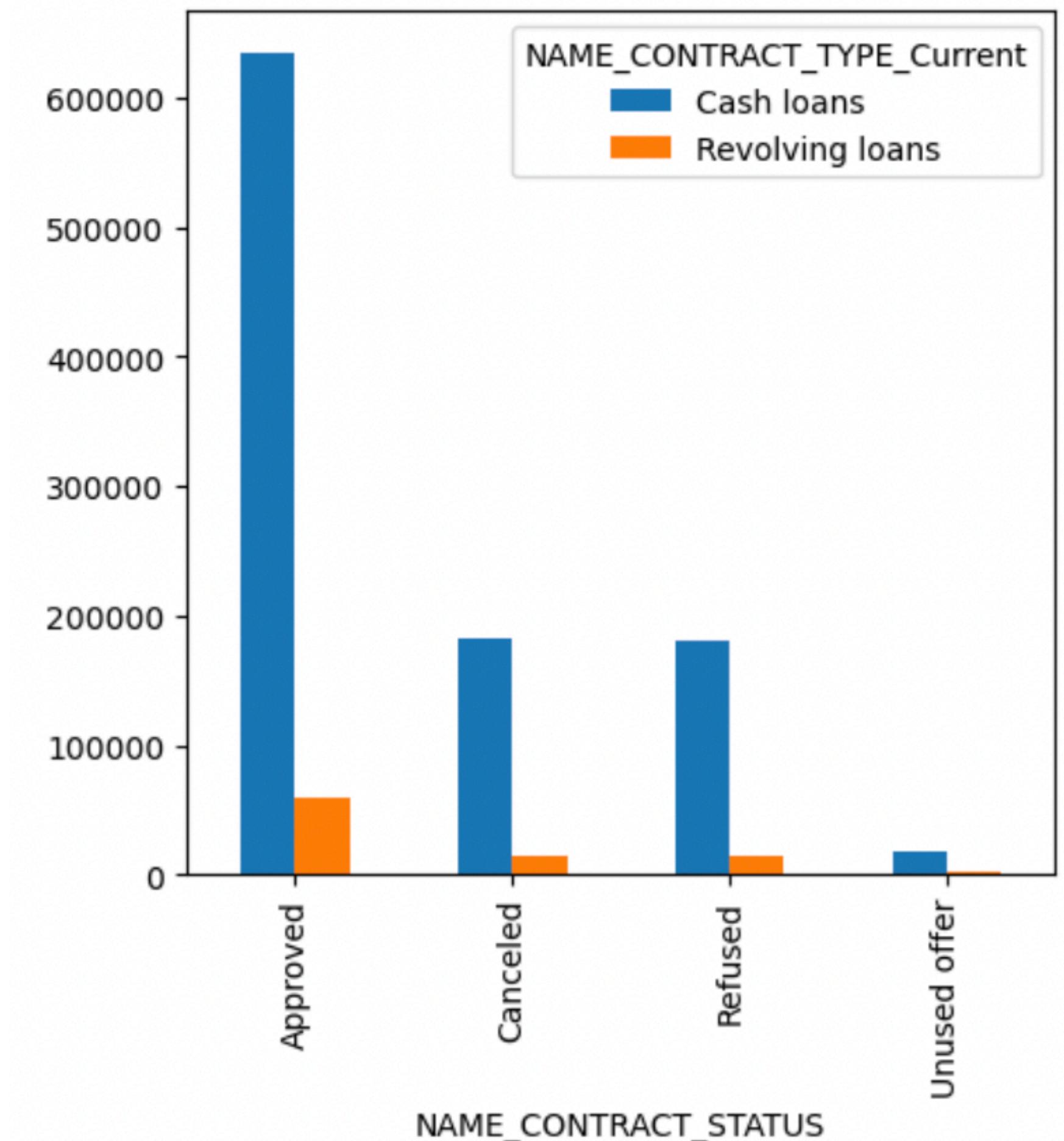
## 1.1 Analysis

"NAME\_CONTRACT\_STATUS" & "NAME\_CONTRACT\_TYPE\_Current"

### Observation:

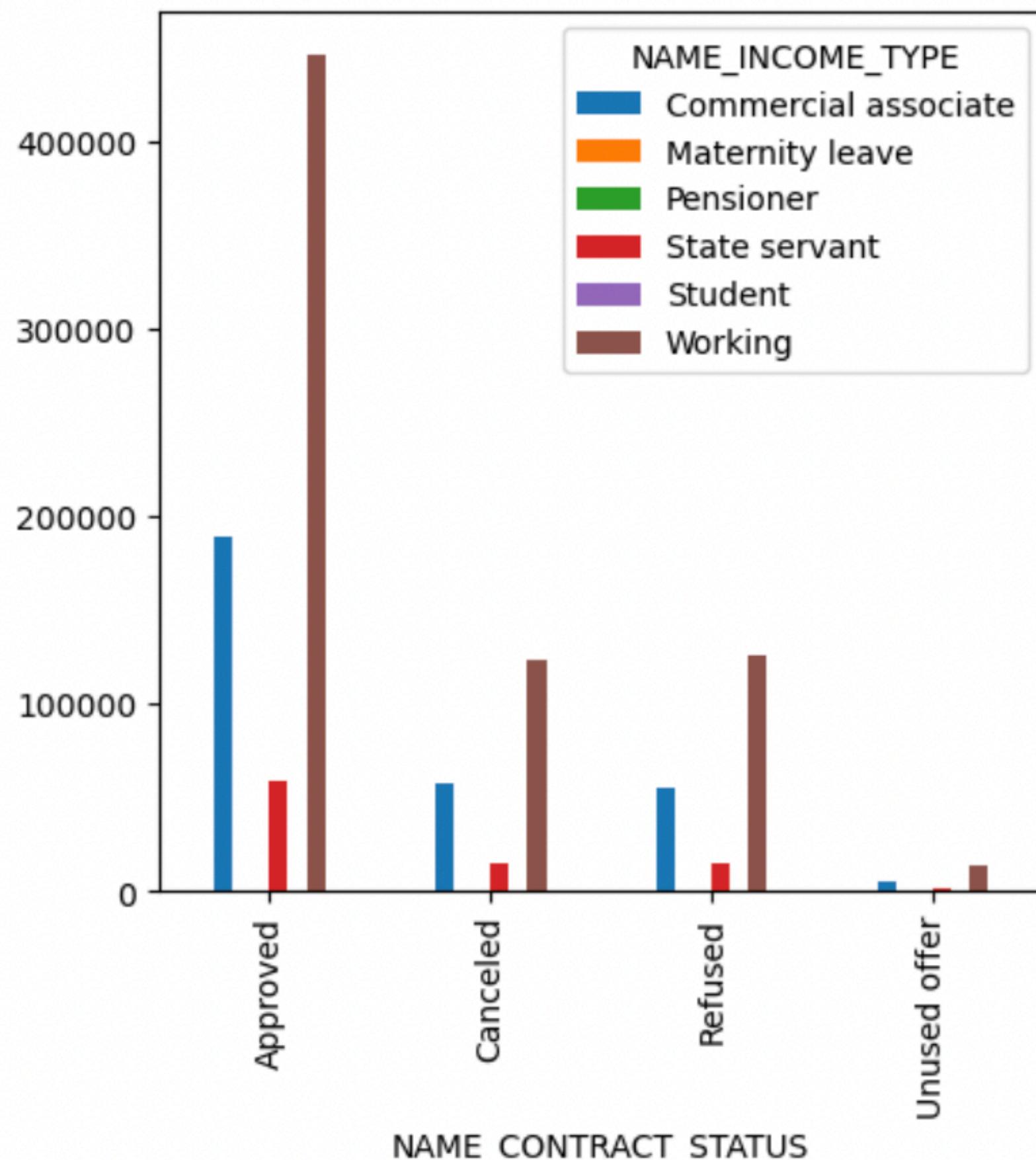
Cash loans have the highest number of approved loans, and overall, the count of cash loans surpasses that of revolving loans in all cases.

NAME_CONTRACT_TYPE_Current	Cash loans	Revolving loans
NAME_CONTRACT_STATUS		
Approved	635143	59601
Canceled	181925	14213
Refused	181590	14440
Unused offer	18299	2023



## 1.2 Analysis “NAME\_CONTRACT\_STATUS” & “NAME\_INCOME\_TYPE”

NAME_INCOME_TYPE	Commercial associate	Maternity leave	Pensioner	State servant	Student	Working
NAME_CONTRACT_STATUS						
Approved	189047	10	44	59205	20	446418
Canceled	57150	2	14	15129	3	123840
Refused	55410	3	26	15062	1	125528
Unused offer	4858	1	0	1465	0	13998



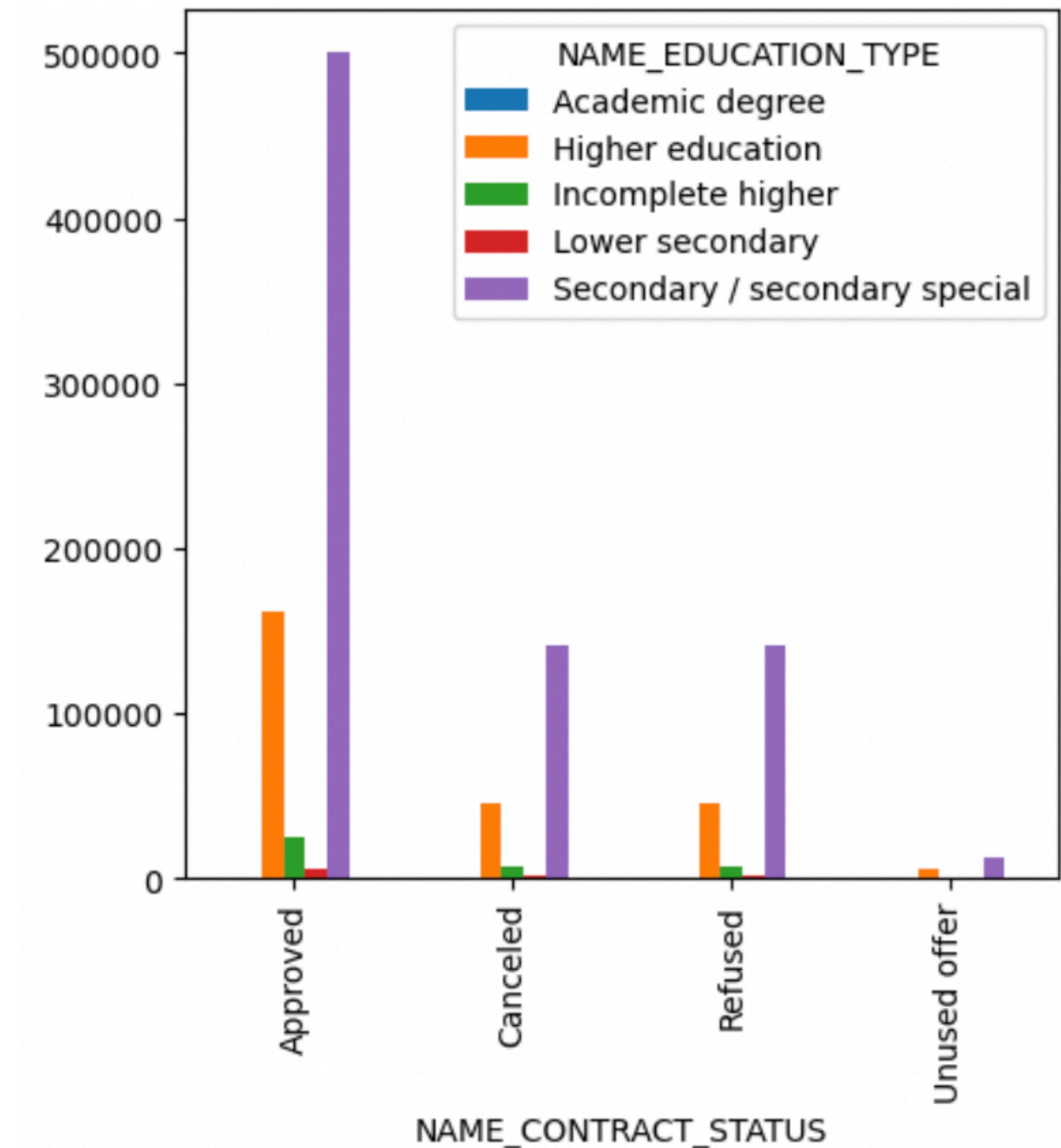
### Observation:

The loan approval rate is highest among working professionals and students compared to other income types.

## 1.3 Analysis “NAME\_CONTRACT\_STATUS” & “NAME\_EDUCATION\_TYPE”

### Observation:

Applicants with a secondary or secondary-special education have the highest rate of approved loans.



---

# **CONCLUSION:**

- **The bank should prioritise clients with higher education levels.**
  - **Loans should be approved for older female clients.**
  - **Clients with previously approved loans should be considered favourable candidates.**
  - **Loans should be granted to clients in higher income categories.**
  - **The bank should be cautious with working professionals, as they have a higher default rate.**
-

---

**THANK YOU!**

---