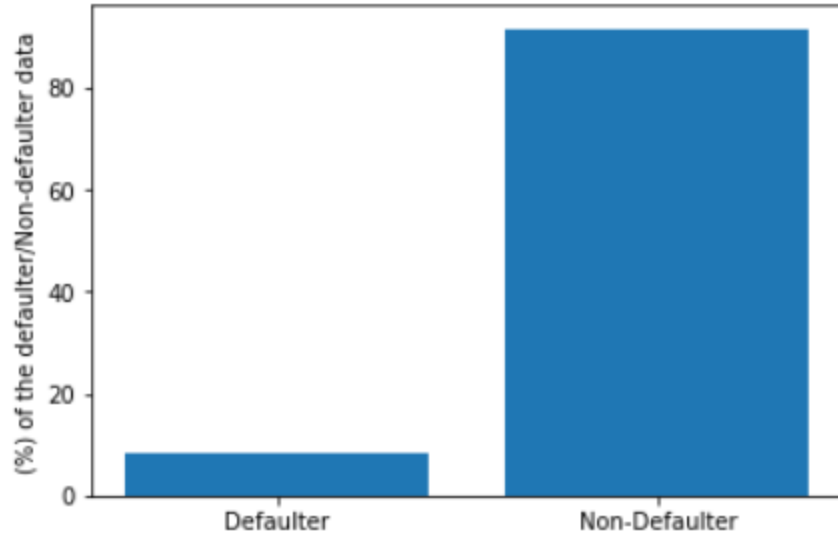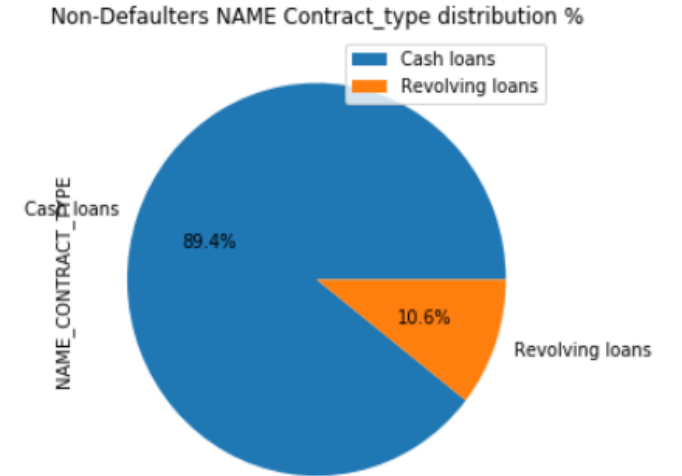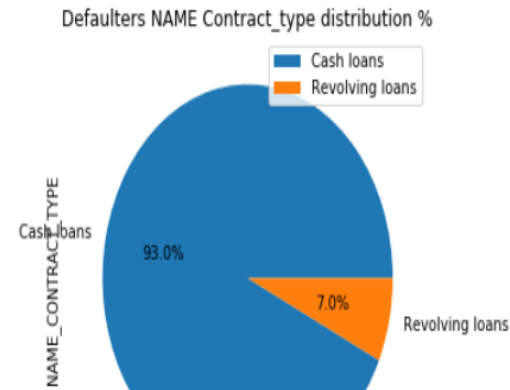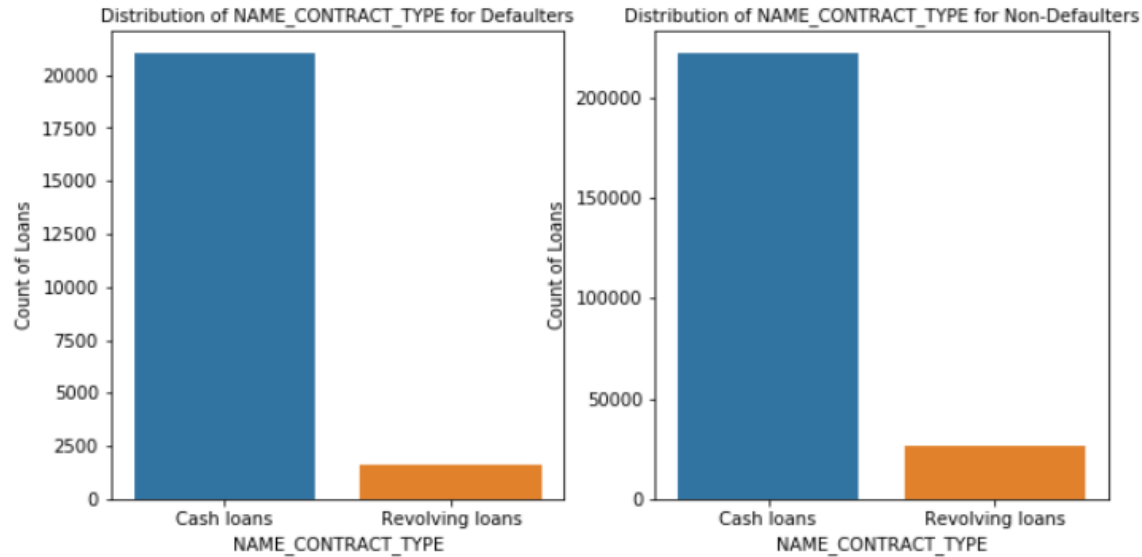# Imbalance percentage

Ratios of imbalance in percentage for non-defaulter and defaulter datas are: 91.646713 and 8.353287
Ratios of imbalance in real-numbers for non-defaulter and defaulter datas is 10.971335 : 1 (approx)
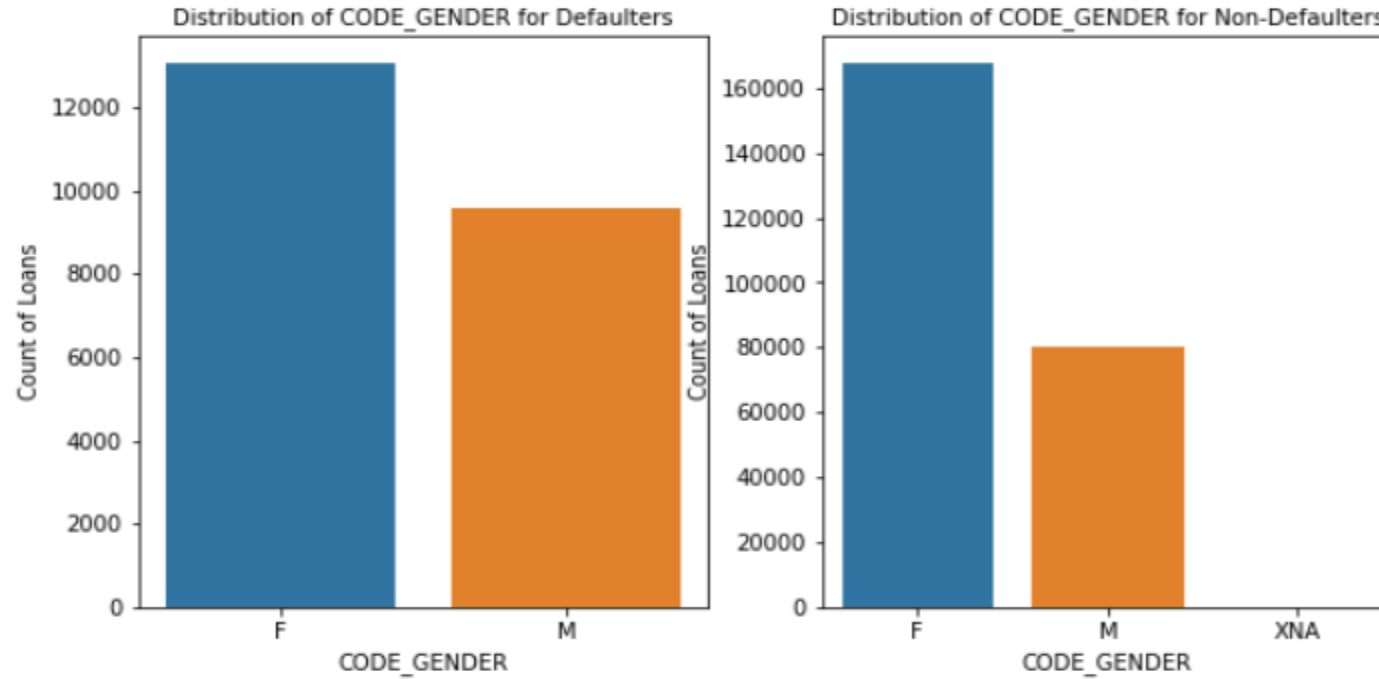
univariate analysis for categorical variables
Application data

Cash loans & Revolving loans

It states that the number of Cash loans is higher than the number of Revolving loans for both Target = 0 and Target = 1.
While distributing cash loans, it is recommended to consider additional back ground check. As the percentage is high for defaulters.

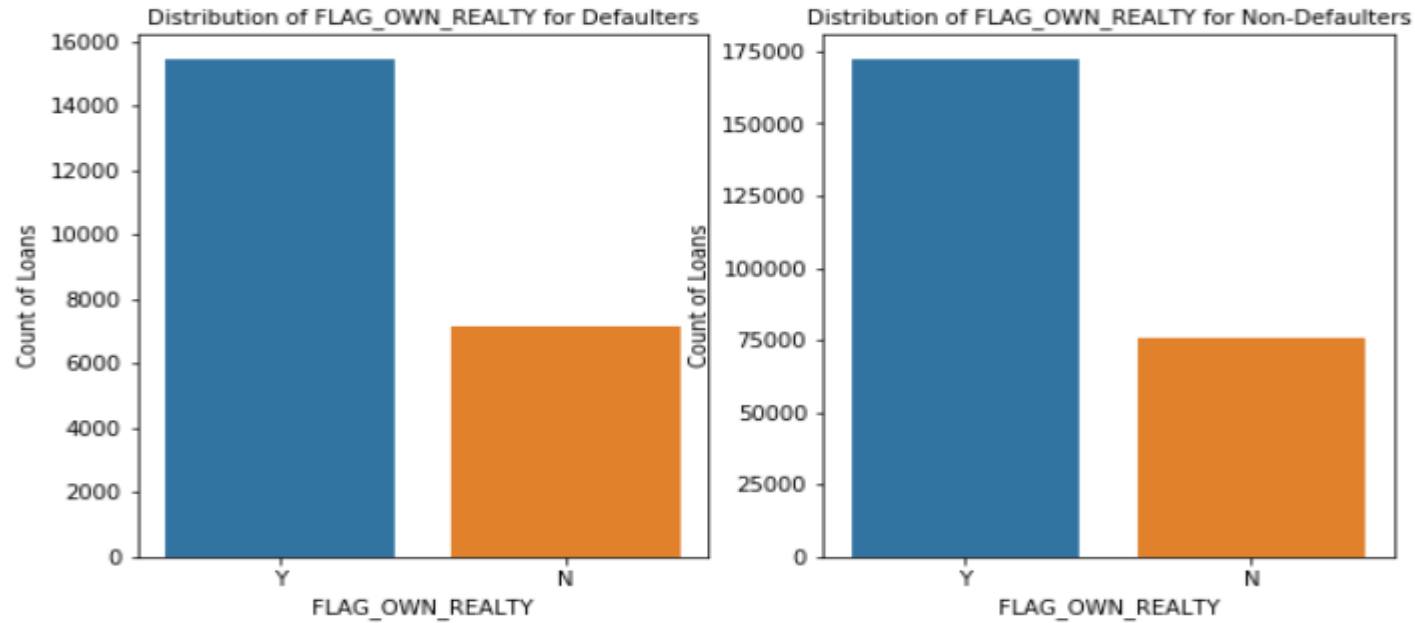# Gender Analysis

It states that the number of Females taking loans is higher than that of Males for both Target = 0 and Target = 1.
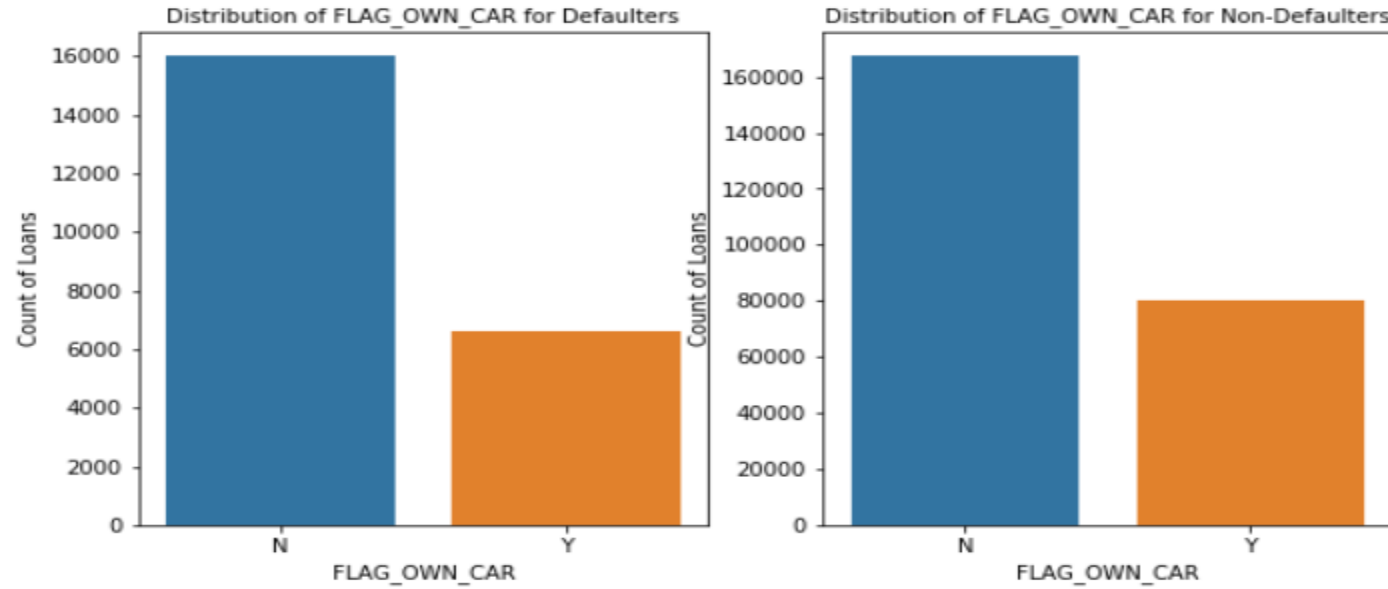
Generally Females take more loans than male. It is more important to concentrate on Female rather than male.
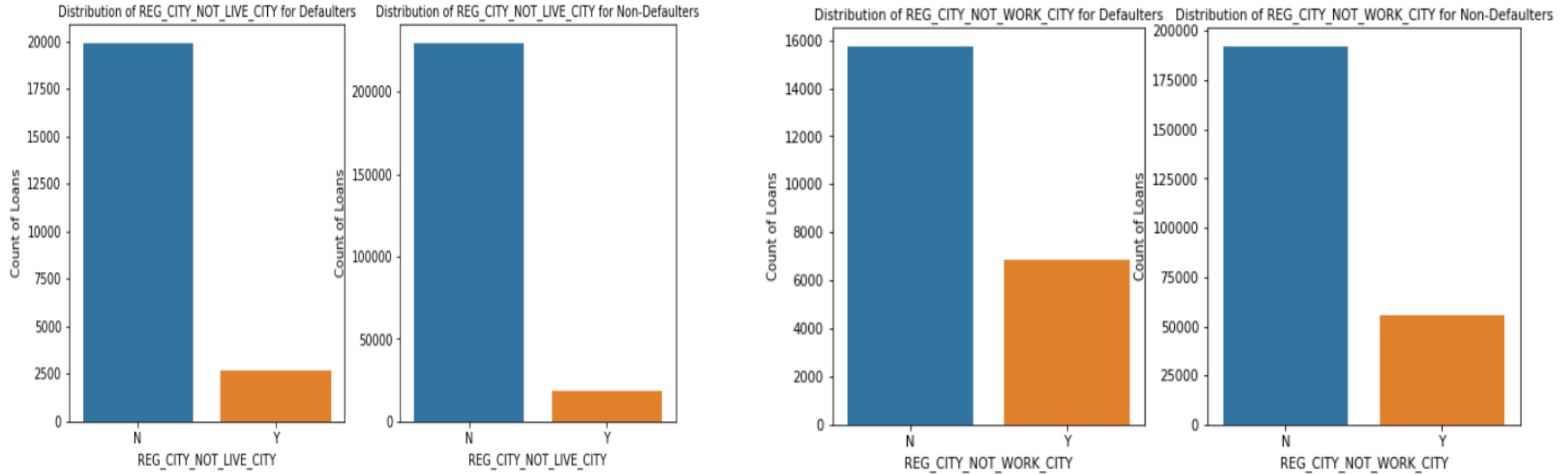
Flag Own Reality

It states that the people who owns a house or a flat tends to opt for loans higher than who don't, also people who tends to own a flat in the Defaulter category are higher to not pay the loan on time.

Flag Own Car

It states that the number of most people applying for loan do not own a car & the ratio of people who own a car is higher for non-defaulters

Live City & Working City

Distribution of REG_CITY_NOT_LIVE_CITY for Defaulters — Distribution of REG_CITY_NOT_LIVE_CITY for Non-Defaulters — Distribution of REG_CITY_NOT_WORK_CITY for Defaulters — Distribution of REG_CITY_NOT_WORK_CITY for Non-Defaulters

It states that the people who live or work in a city different than the registration city are more likely to become defaulters
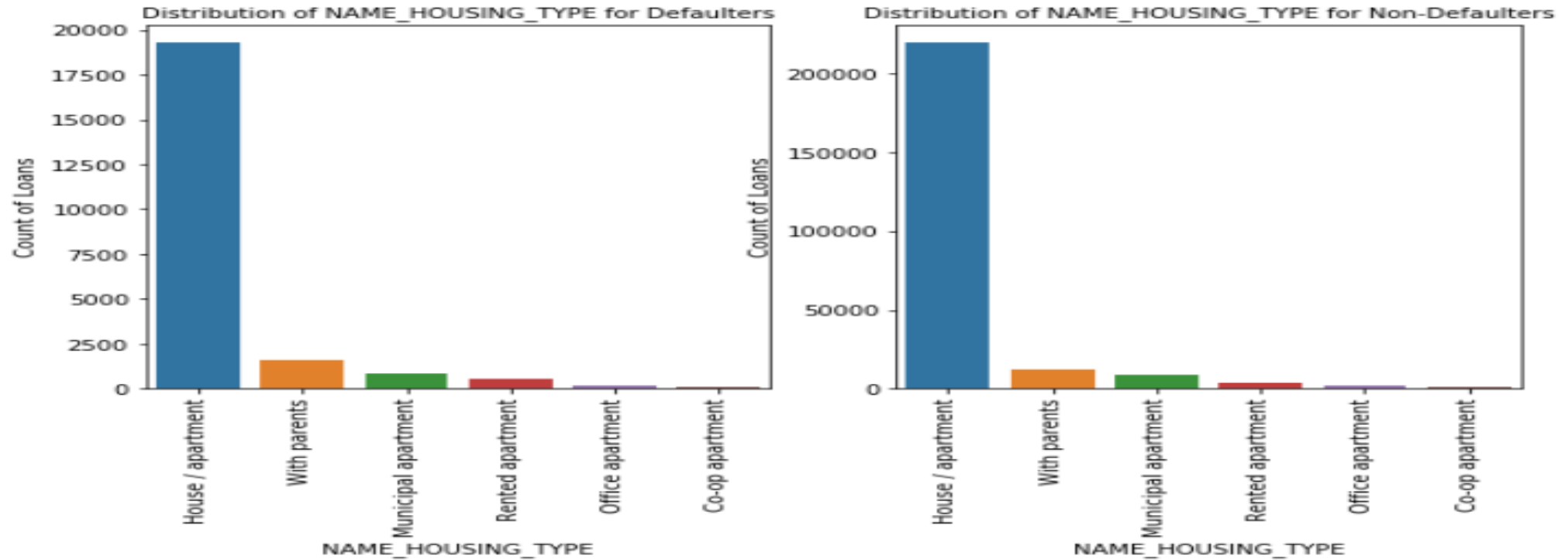
Family Status
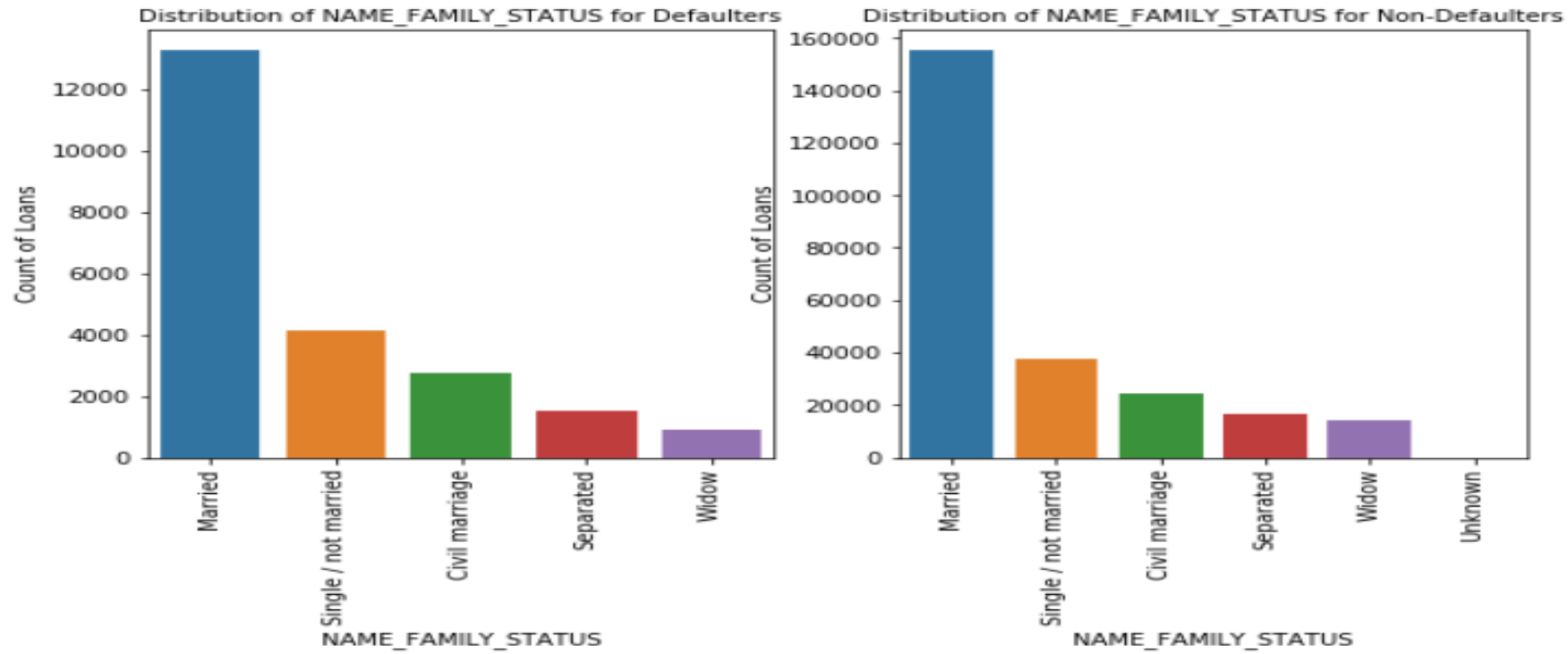
It states that Married people are more likely to opt for loans and Single/Unmarried people are more likely to be defaulters

In case of Single/not Married, 10% of loans are default when compared with Defaulter & Non-Defaulters ratio

Housing Type

It states that people who live in a House/Apartment are most likely to opt for loans.

Ratio of People who live with Parents is more for defaulter than non-defaulters. It says that applicant who live with parents have a higher chance of having payment difficulties.
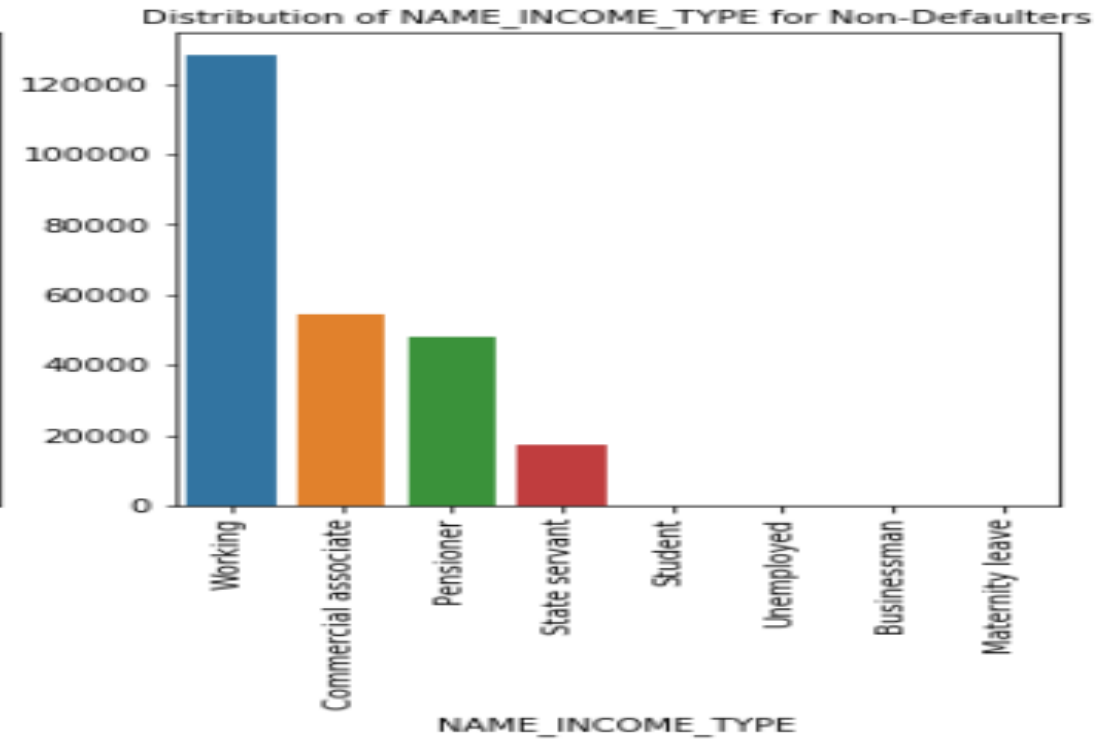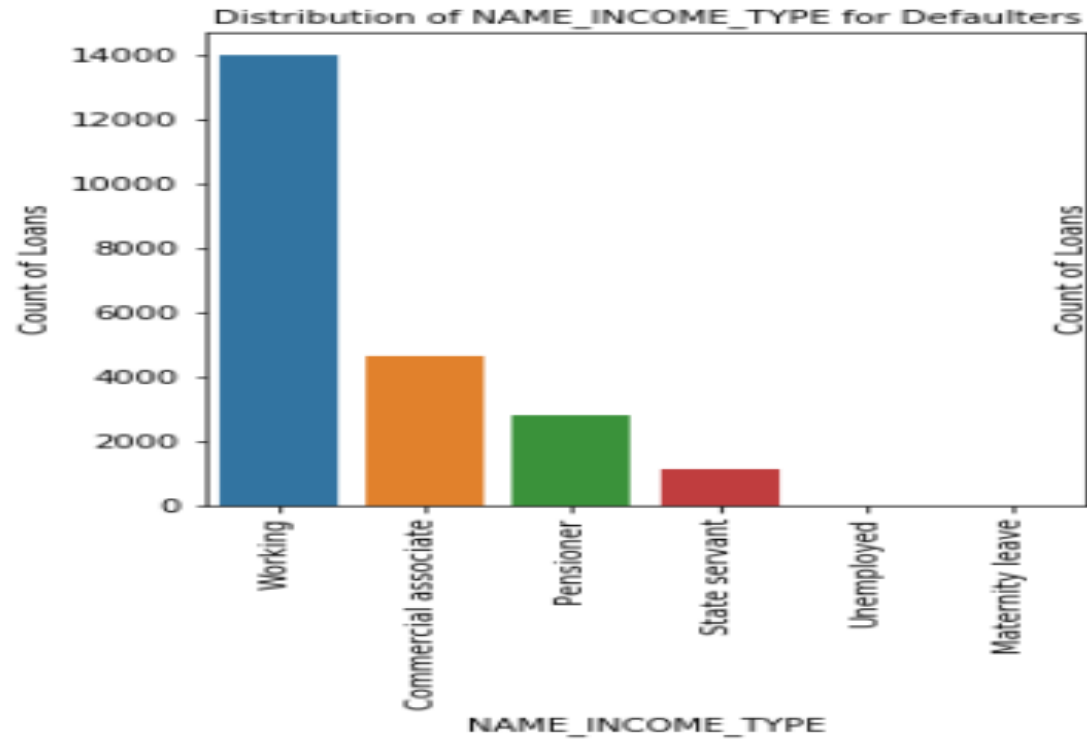
Education Type

Distribution of NAME_EDUCATION_TYPE for Defaulters

Distribution of NAME_EDUCATION_TYPE for Non-Defaulters
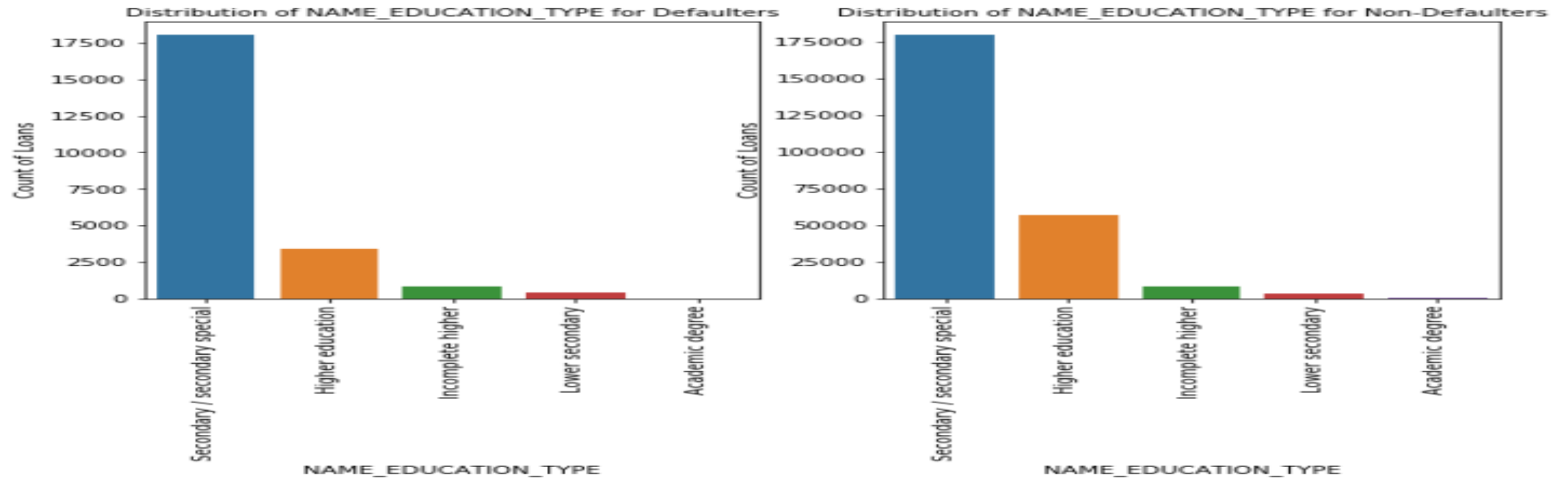
The chart states that people with Academic Degree rarely take loans and are rarely defaulters. So they are potentially good customers.

People with higher education are less likely to have payment difficulties. The Ratio is higher for non-defaulters than defaulters.
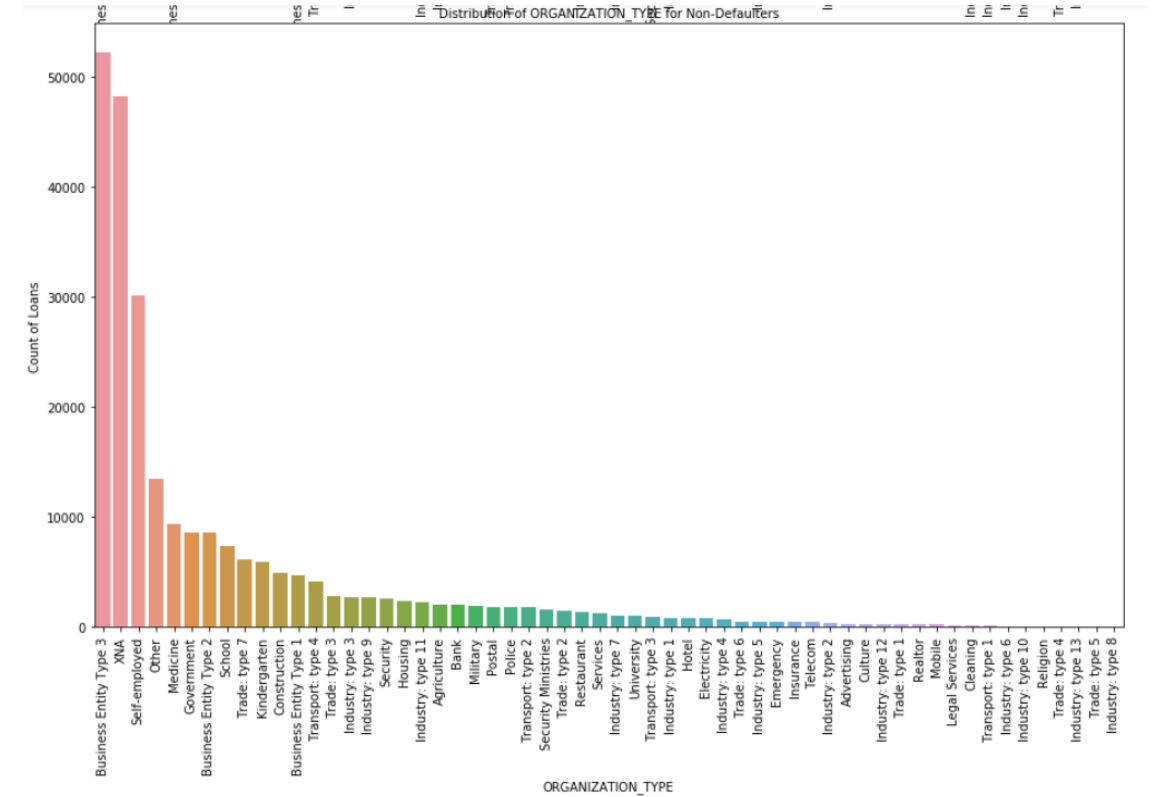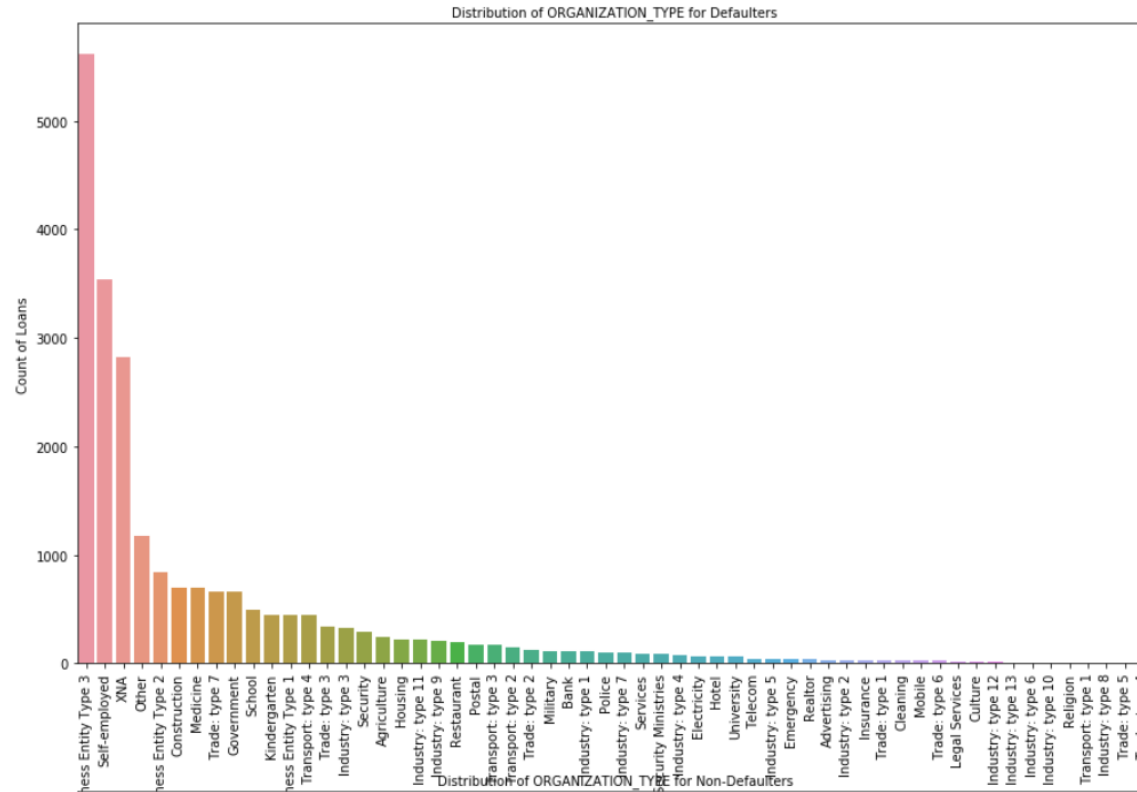
Income Type

It states Commercial associates, Pensioner, State Servants have a higher ratio to total in non-defaulters.

Working class is the largest loan takers and highest in defaulting loans.

Organization Type

Distribution of ORGANIZATION_TYPE for Defaulters

Distribution of ORGANIZATION_TYPE for Non-Defaulters

It states that Business Entity Type 3, XNA & Self-Employed take the highest no. of loans and makes the majority of defaulters.
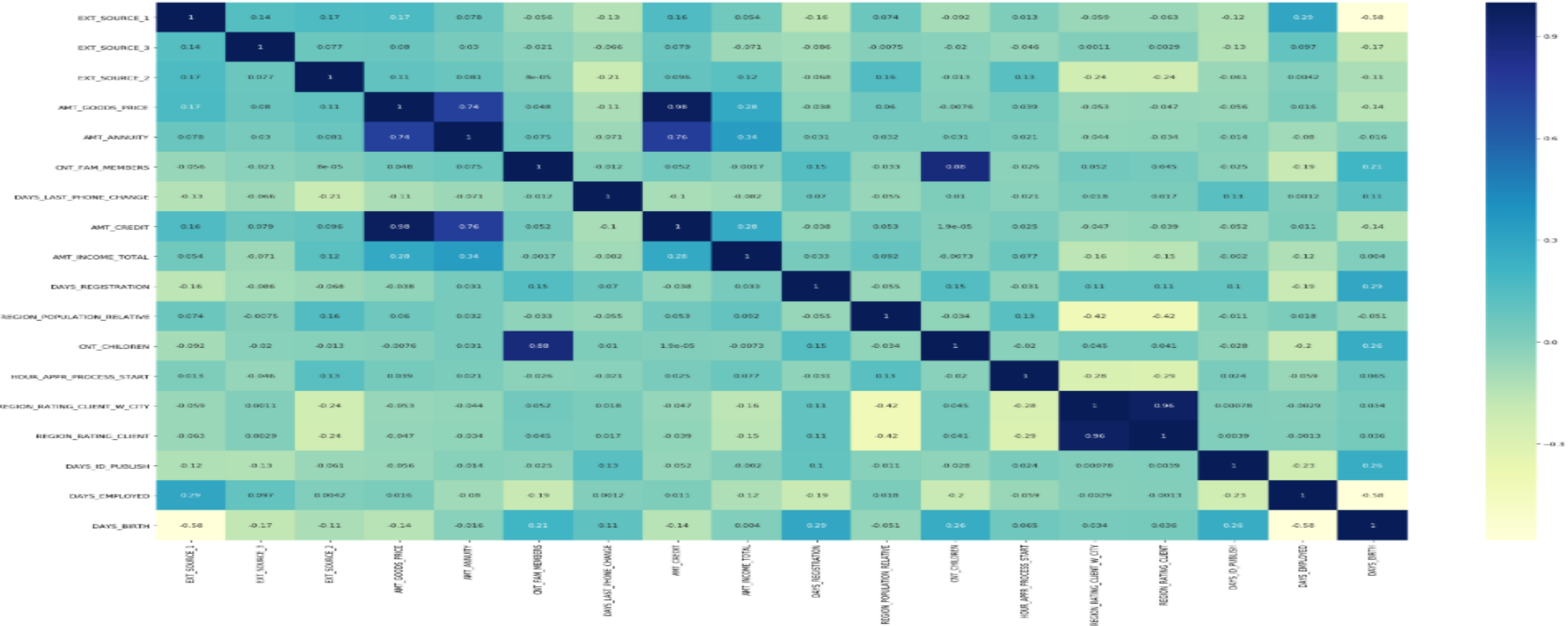
Income Bracket
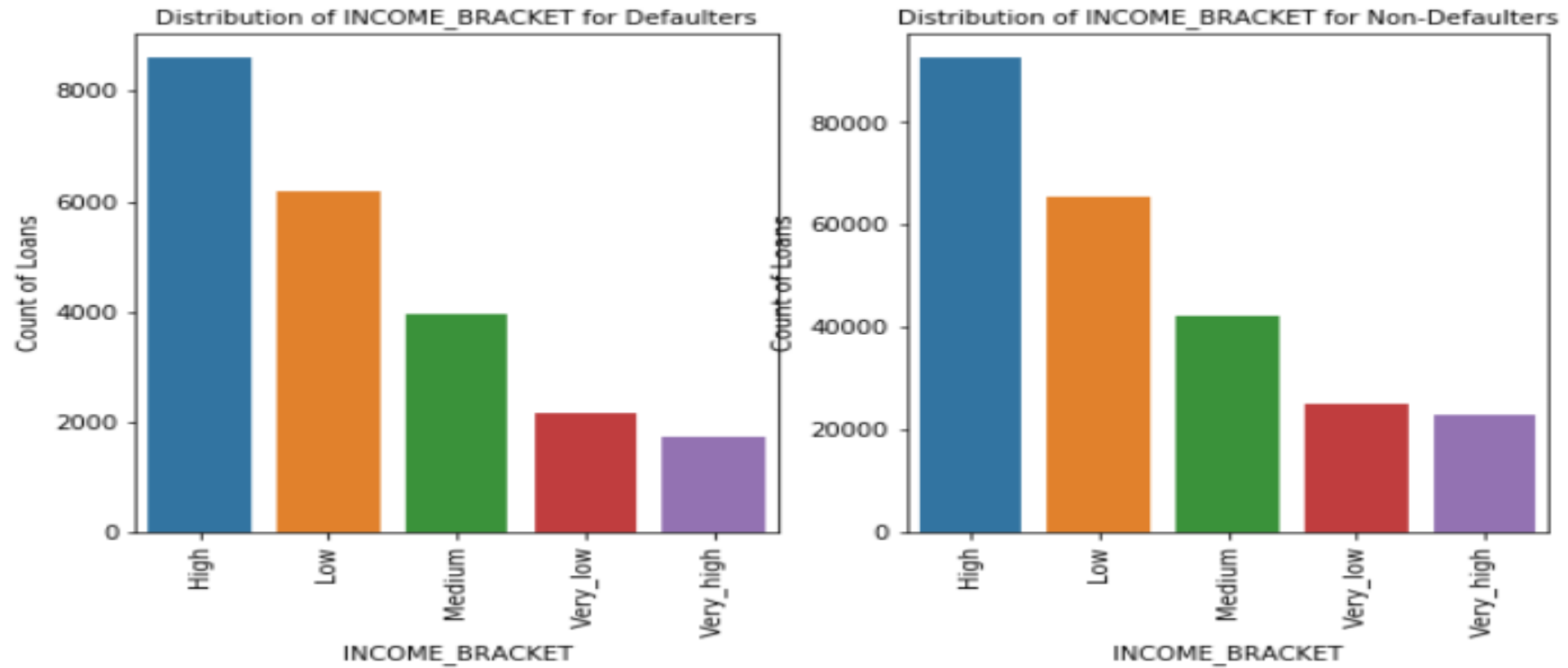
It states that the people with High Income Brackets where the majority of loan defaulters fall.

Bi-Variate Analysis of Variables & Correlation Application data

Correlation for Defaulters

Top positive correlations:

* AMT_CREDIT - AMT_GOODS_PRICE

* REGION_RATING_CLIENT_W_CITY - REGION_RATING_CLIENT

* CNT_CHILDREN - CNT_FAM_MEMBERS

* AMT_CREDIT - AMT_ANNUITY

* AMT_GOODS_PRICE - AMT_ANNUITY

Top negative correlations:

* HOUR_APPR_PROCESS_START - REGION_RATING_CLIENT_W_CITY

* REGION_RATING_CLIENT - HOUR_APPR_PROCESS_START

* REGION_POPULATION_RELATIVE - REGION_RATING_CLIENT

* REGION_RATING_CLIENT_W_CITY - REGION_POPULATION_RELATIVE

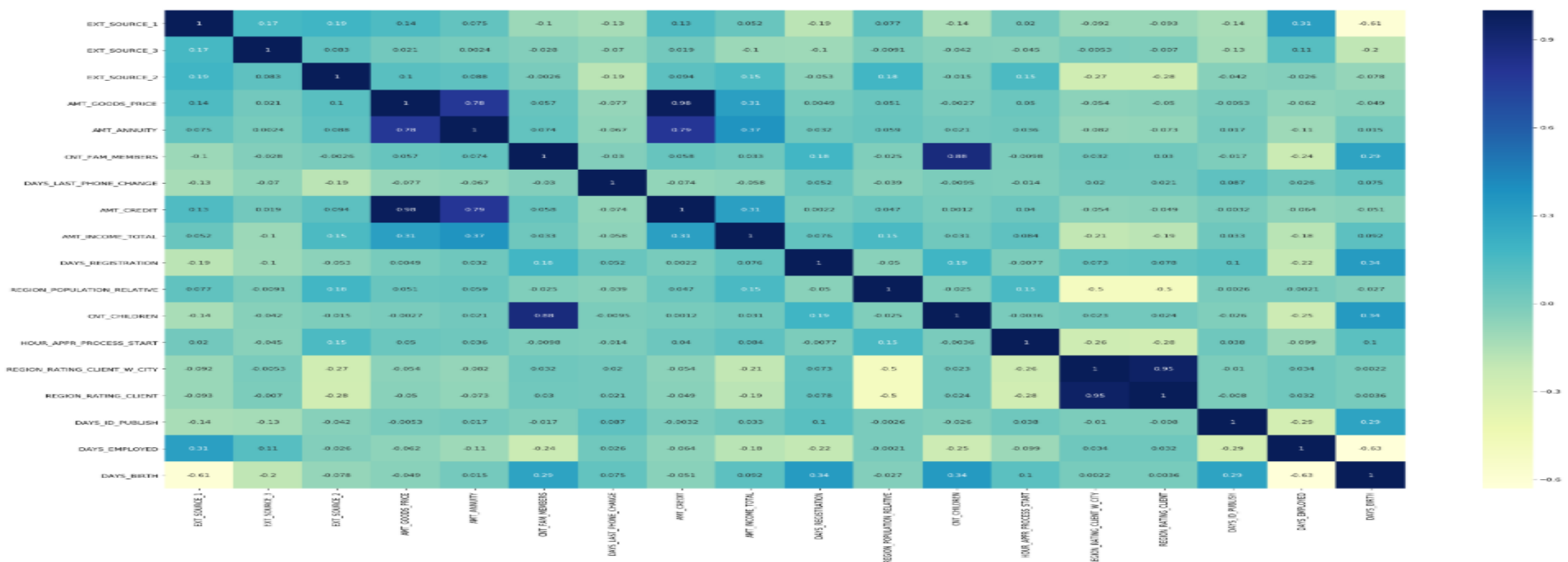* EXT_SOURCE_1 - DAYS_BIRTH

# Correlation for Non-Defaulters

Top positive correlations:

* AMT_CREDIT - AMT_GOODS_PRICE

* REGION_RATING_CLIENT_W_CITY - REGION_RATING_CLIENT

* CNT_CHILDREN - CNT_FAM_MEMBERS

* AMT_CREDIT - AMT_ANNUITY

* AMT_GOODS_PRICE - AMT_ANNUITY

Top negative correlations:

* HOUR_APPR_PROCESS_START - REGION_RATING_CLIENT_W_CITY

* REGION_RATING_CLIENT - HOUR_APPR_PROCESS_START

* REGION_POPULATION_RELATIVE - REGION_RATING_CLIENT

* REGION_RATING_CLIENT_W_CITY - REGION_POPULATION_RELATIVE
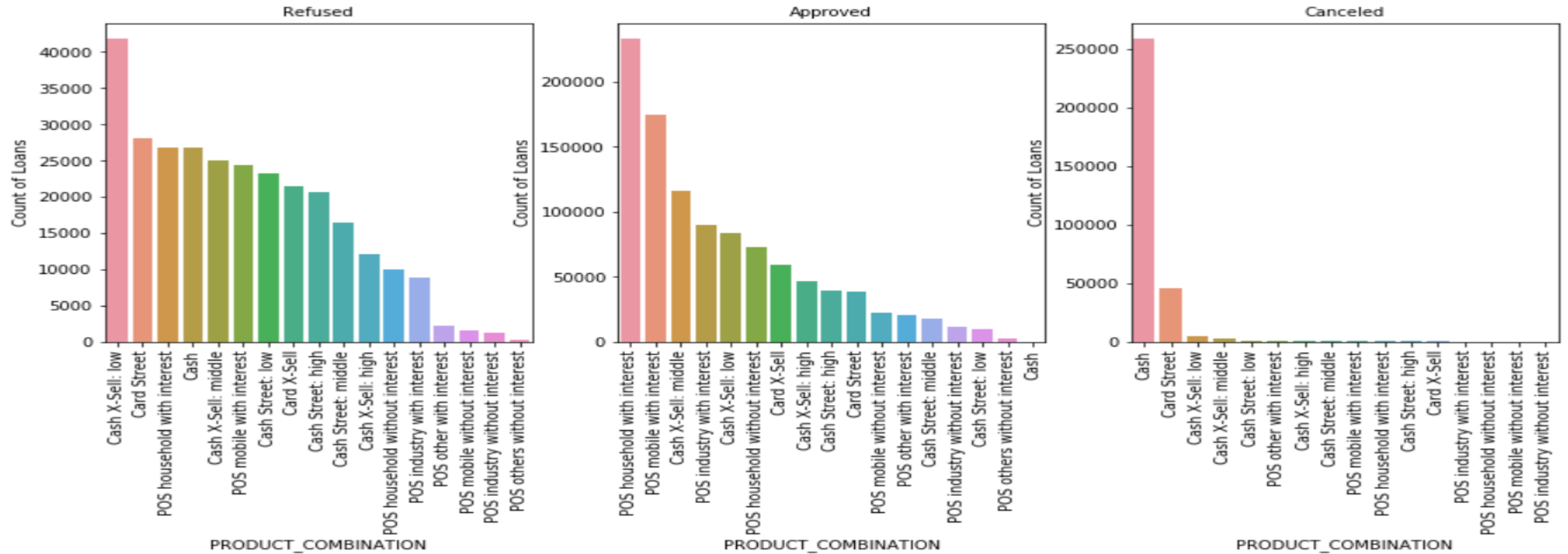
* EXT_SOURCE_1 - DAYS_BIRTH

univariate analysis for categorical variables
Previous data

Contract Status
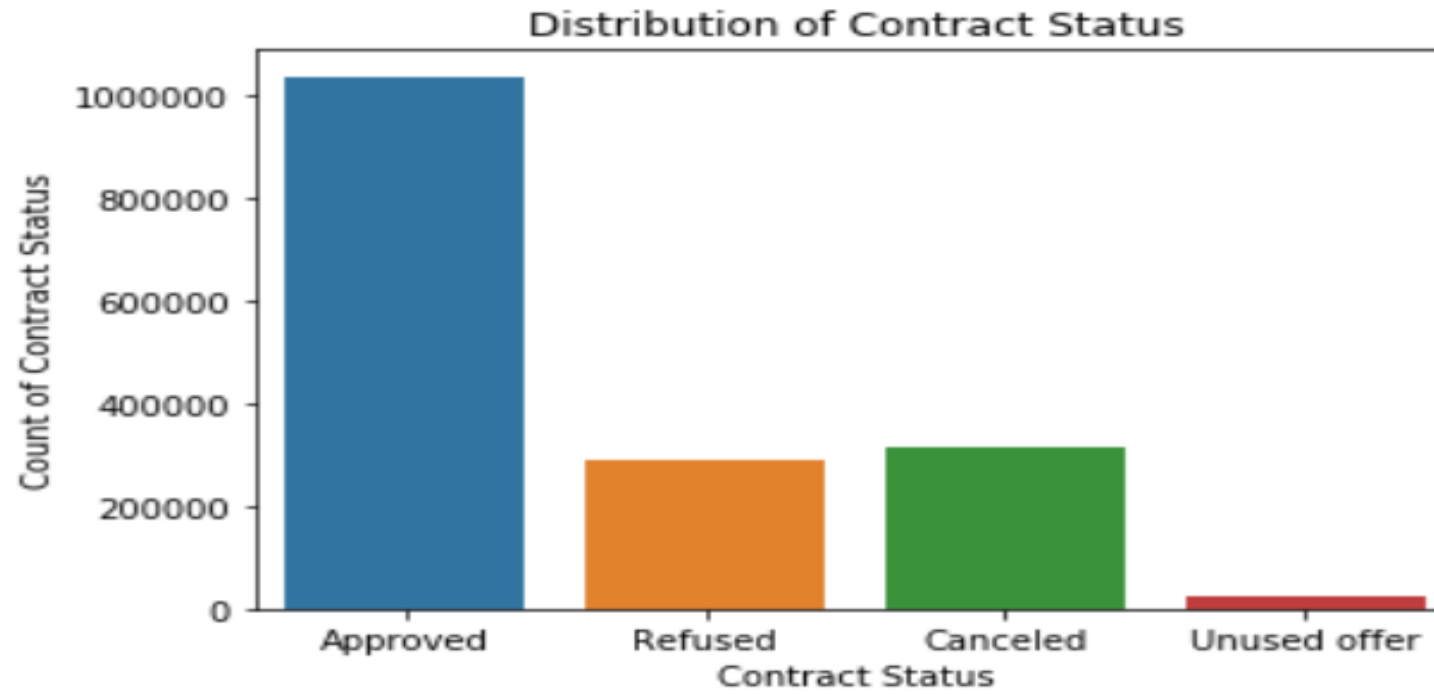
It states that the people with High Income Brackets where the majority of loan defaulters fall.
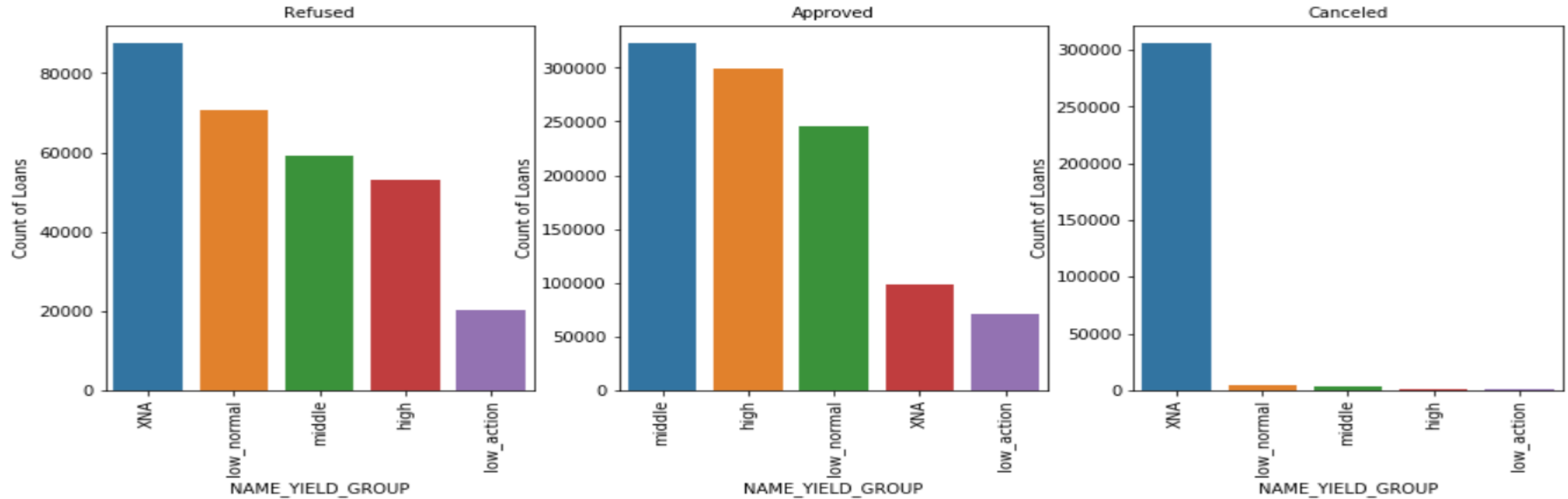
Product Combination

It states that most number of loans were approved for POS household with interest.

Most number of refused loans were of Cash X-Sell: Low Product combination.

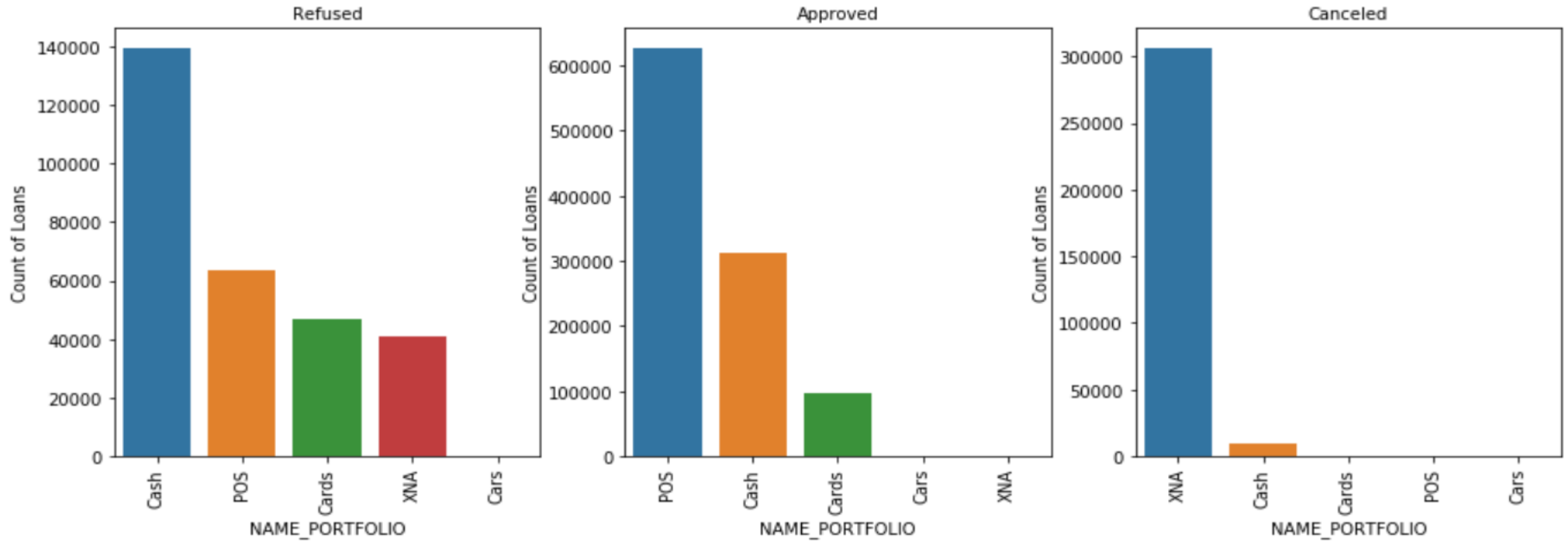Most Canceled loans were Cash loans.

Name Yield Group

It states that most approved loans were from Middle Yield Goup.

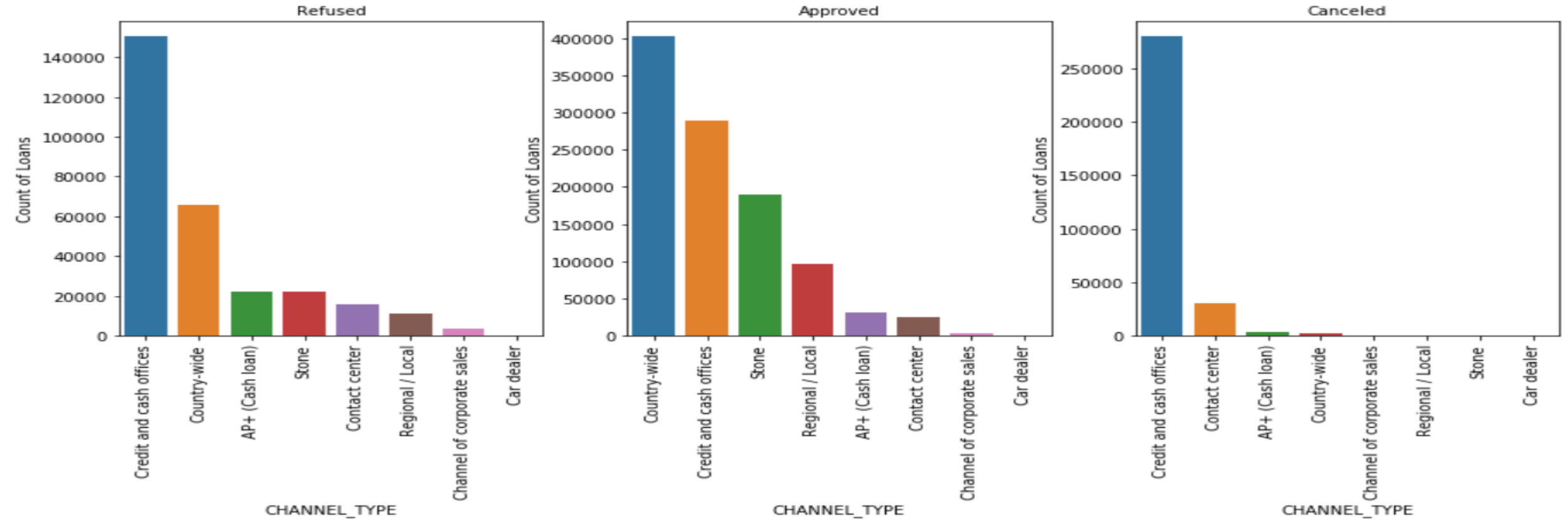**Most refused loans were from Yield Goups Not specified.**

Name Portfolio

It states that most approved loans were POS.

Most refused loans were Cash.

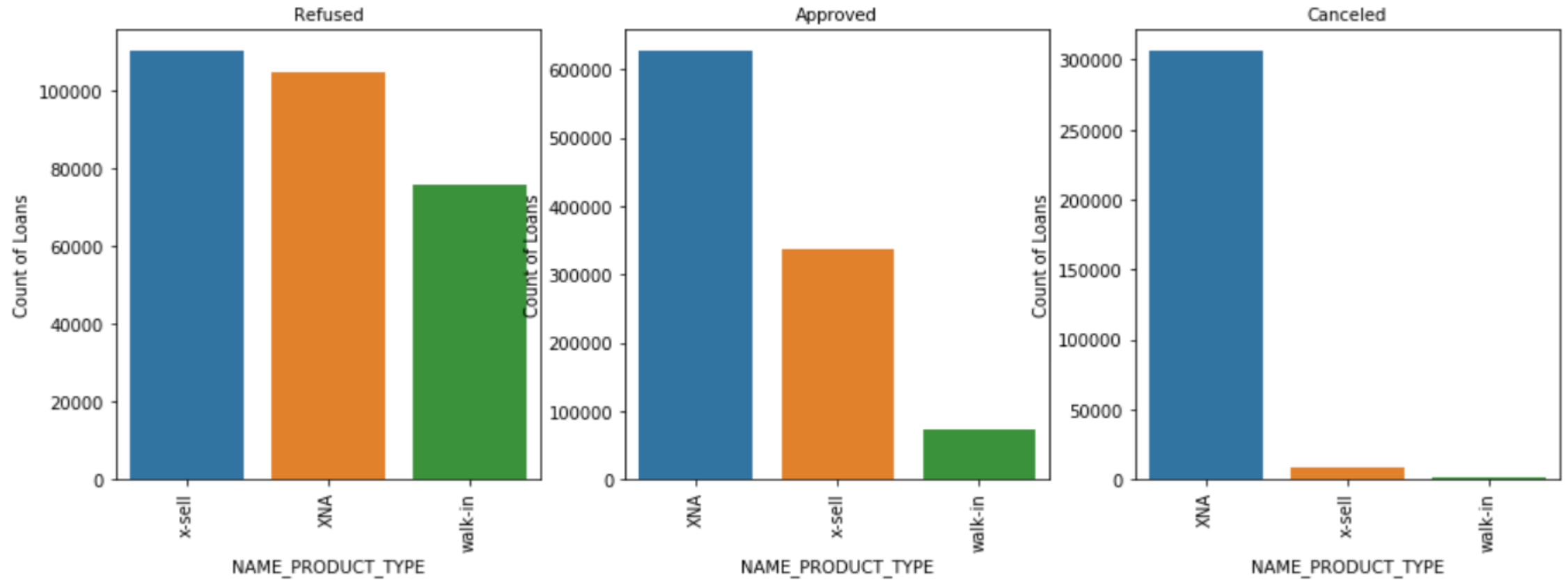# Channel Type

It states that most approved loans were from Country-wide Channel.

Most refused loans were from Credit and Cash Offices Channel.
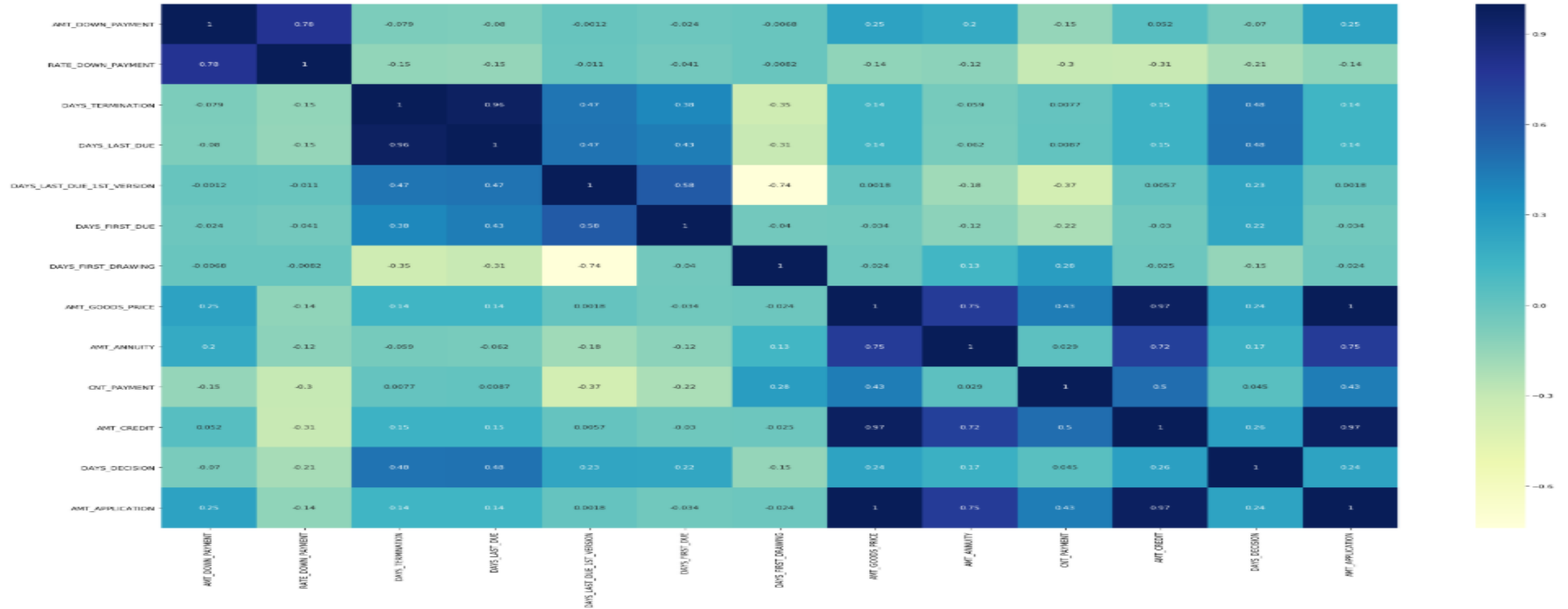
Name Product Type

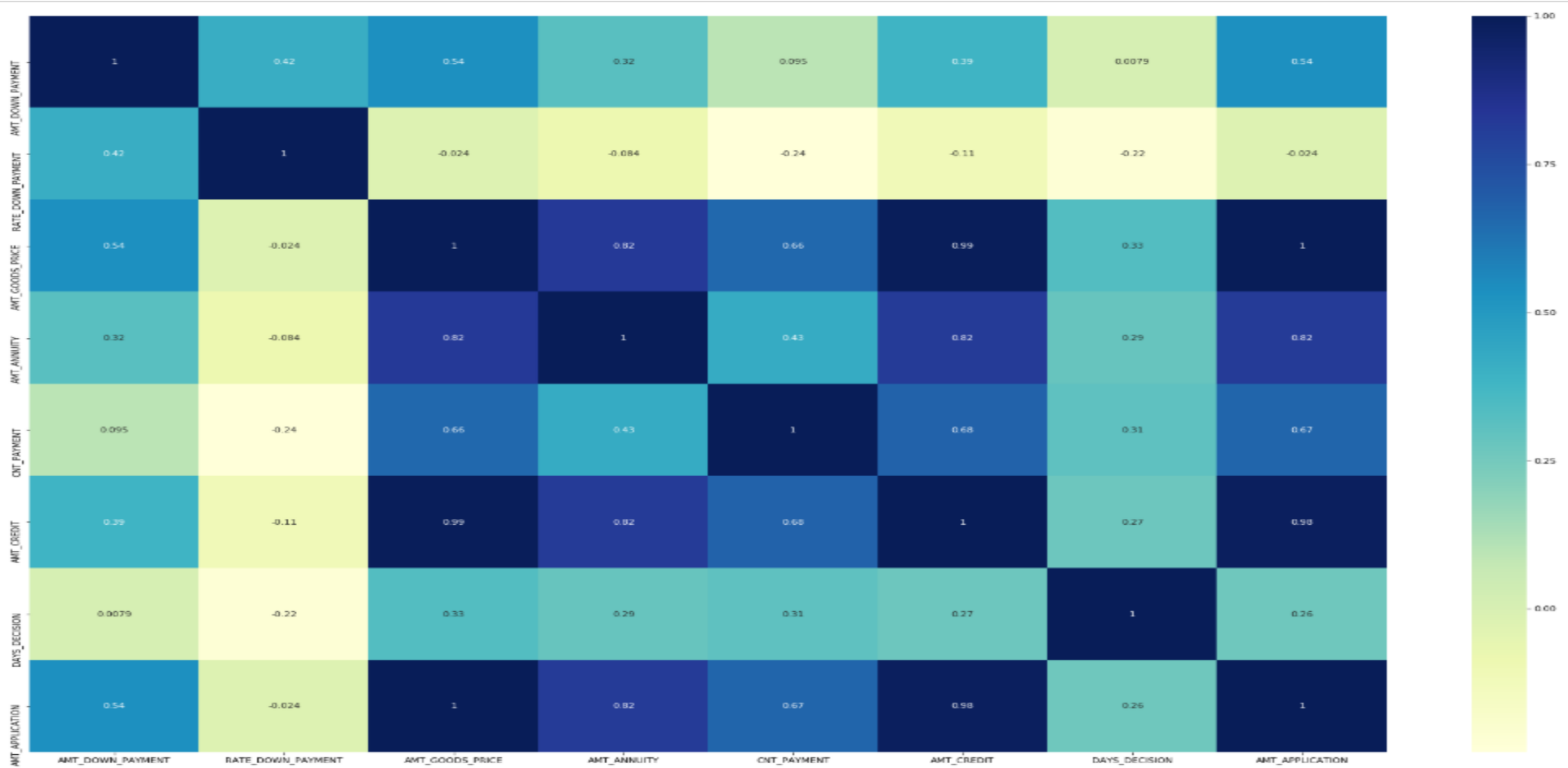It states that most approved loans were from Country-wide Channel.

Most refused loans were from Credit and Cash Offices Channel.

Bivariate Analysis for Previous application data

Correlation

Correlation

Univariate Analysis for Application & Previous application data

Previously Not Approved Defaulters
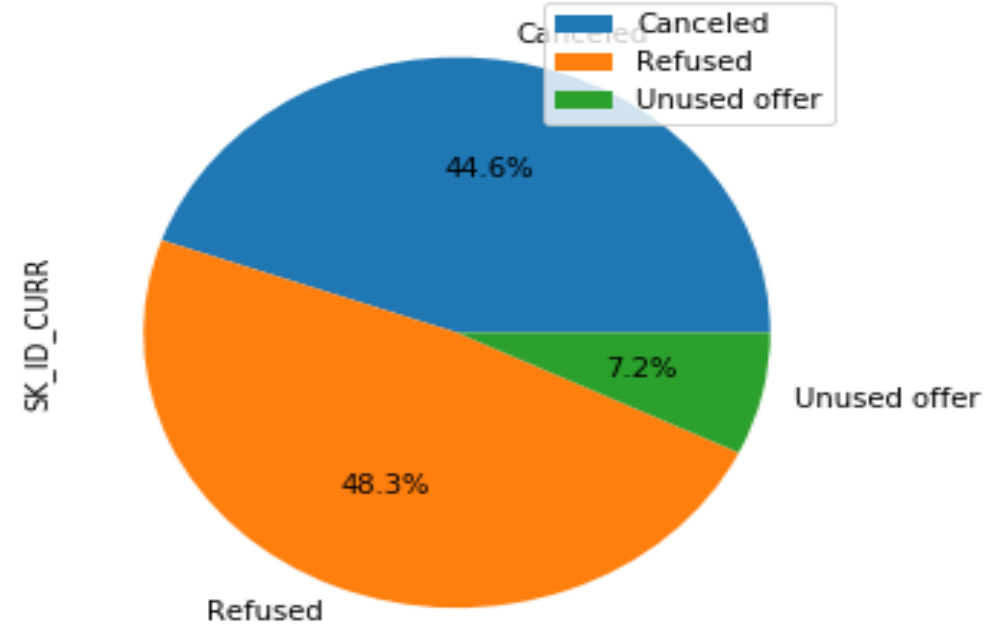
48.3%(9441) of the loan applications refused previously were facing the difficulty in achieving target

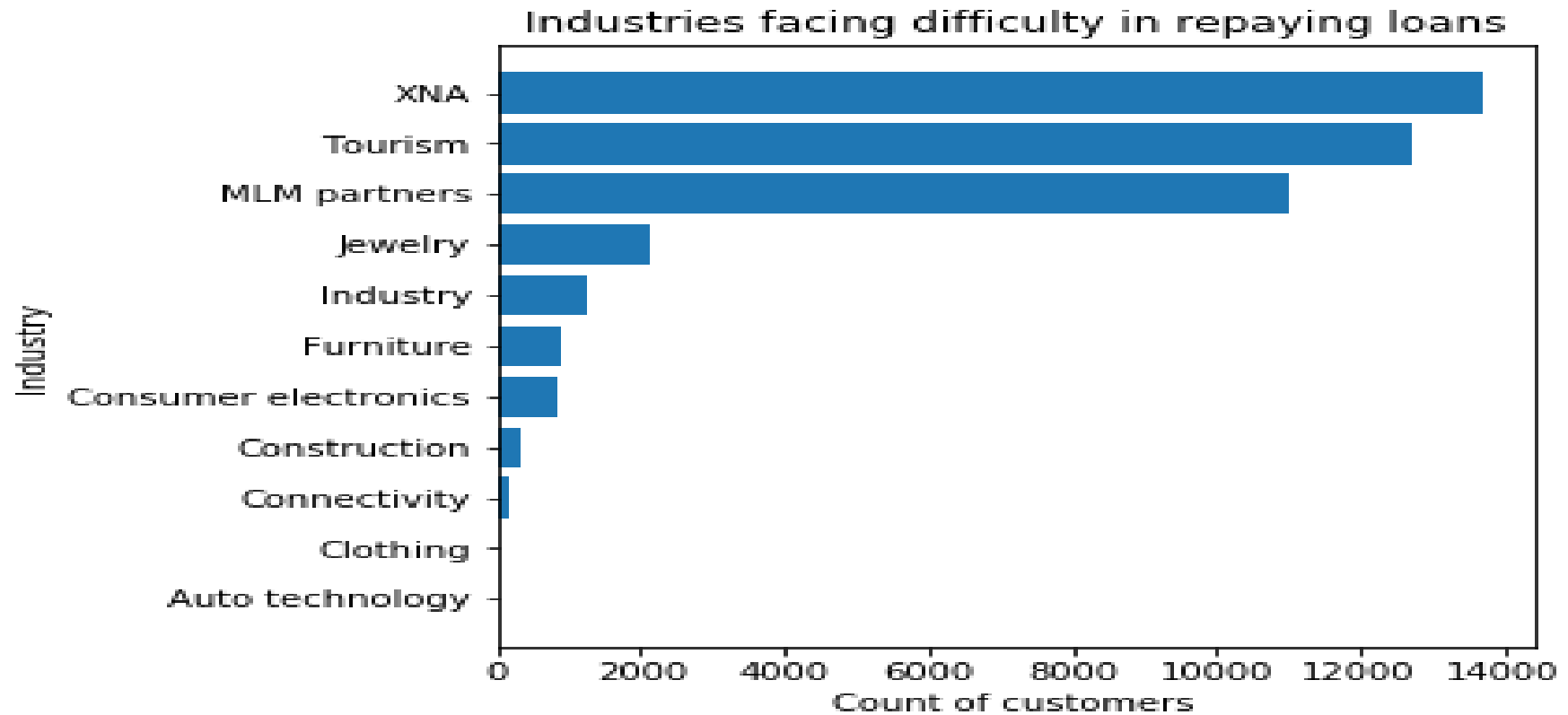**44.6%(8716) of the loan applications cancelled previously were facing the difficulty in achieving target**

7.2%(1405) of the loan applications unused offer were facing the difficulty in achieving target

Previous Industries defaulters

Industries facing difficulty in repaying loans
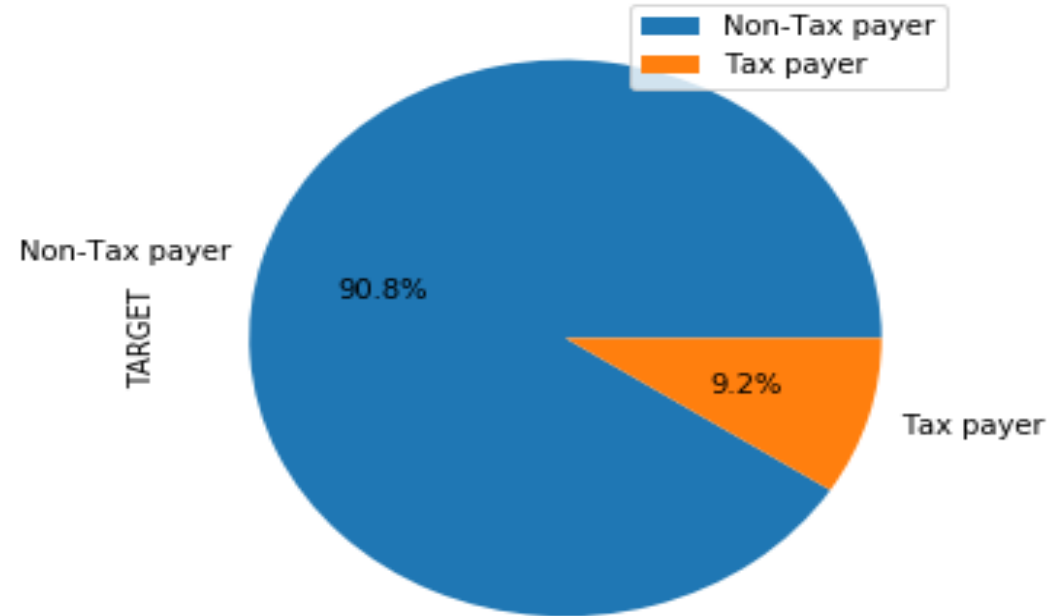
Industries such as Tourism & MLM Partners are facing difficulty in repaying loans.

Derived metrics

Income Group - Defaulters

The income group of Non-Tax payer has more defaulters where as the ratio of Non-Tax payer and Tax payer is 10:1

Recommendations

- **Application Data - Univariate Analysis:**

Based on our analysis, following are the key drivers for defaulters.
  - Own Car
  - Live City & Working City
  - Income Type
  - Family Status
  - Income Bracket

Recommendations

- **Application Data - Bi Variate Analysis:**

Based on our analysis, top positive correlations are:

- AMT_CREDIT - AMT_GOODS_PRICE
- REGION_RATING_CLIENT_W_CITY - REGION_RATING_CLIENT
- CNT_CHILDREN - CNT_FAM_MEMBERS
- AMT_CREDIT - AMT_ANNUITY
- AMT_GOODS_PRICE - AMT_ANNUITY

Based on our analysis, top negative correlations:

- HOUR_APPR_PROCESS_START - REGION_RATING_CLIENT_W_CITY
- REGION_RATING_CLIENT - HOUR_APPR_PROCESS_START
- REGION_POPULATION_RELATIVE - REGION_RATING_CLIENT
- REGION_RATING_CLIENT_W_CITY - REGION_POPULATION_RELATIVE
- EXT_SOURCE_1 - DAYS_BIRTH

Recommendations

Application & Previous Data:

- Previously Loan Not Approved but Defaulters as per Application Data
- Previous Industries defaulters

Derived Metrics:

- The income group of Non-Tax payer has more defaulters where as the ratio of Non-Tax payer and Tax payer is 10:1