Loan Credit Risk Analysis

Ganesh Jalakam

1. Problem Statement

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specializes in lending various types of loans to urban customers. You must use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.
When the company receives a loan application, the company must decide for loan approval based on the applicant's profile.

Two types of risks are associated with the bank's decision:

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
2. If the applicant is not likely to repay the loan, i.e., he/she is likely to default, then approving the loan may lead to a financial loss for the company.

The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

1. **The client with payment difficulties:** he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample.
2. **All other cases:** All other cases when the payment is paid on time.

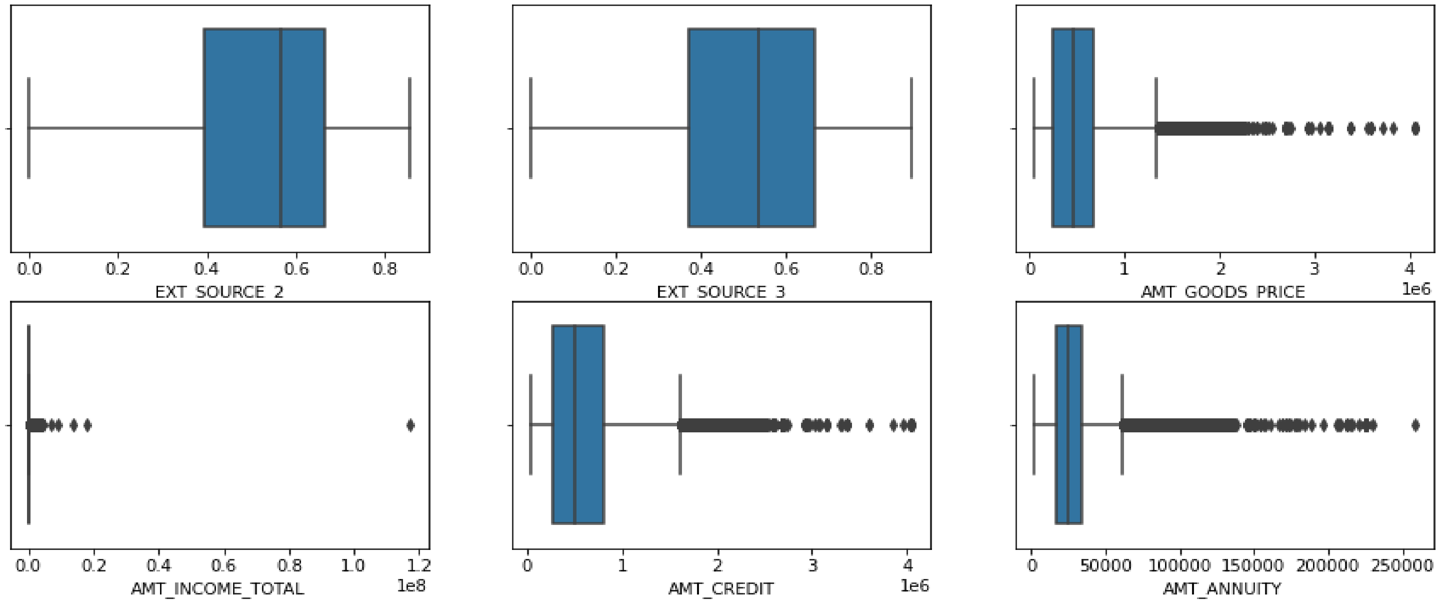
When a client applies for a loan, there are four types of decisions that could be taken by the client/company:

- **Approved**: The Company has approved loan Application
- **Cancelled:** The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
- **Refused:** The company had rejected the loan (because the client does not meet their requirements etc.).
- **Unused offer:** Loan has been cancelled by the client but on different stages of the process.

In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

2. Current Loan Application Data Analysis

2.1 Handling Outliers



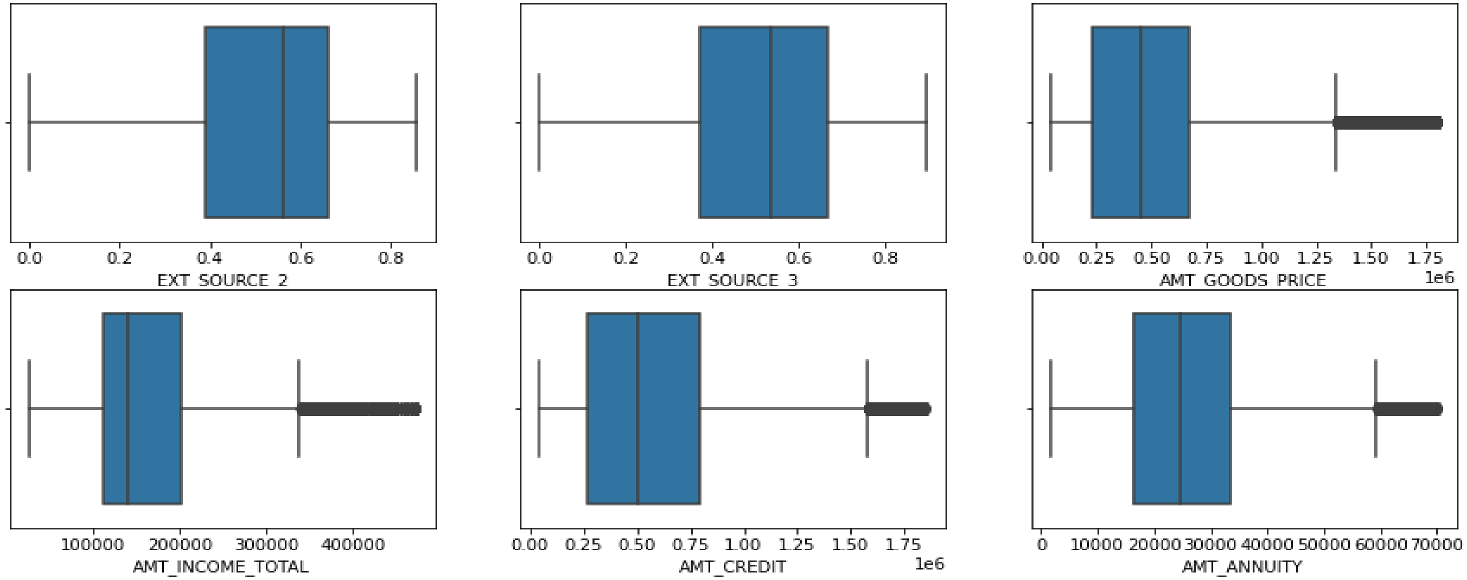
Observation:

From the above plots, we can see the outliers for the following variables:

1.AMT_GOODS_PRICE
2.AMT_INCOME_TOTAL
3.AMT_CREDIT
4.AMT_ANNUITY

We can clearly state that there is a huge difference between 99th percentile and max value for **AMT_GOODS_PRICE, AMT_INCOME_TOTAL, AMT_CREDIT, AMT_ANNUITY** variables. We can safely drop the records which falls outside of 99th percentile for these variables.

After Treating Outliers:

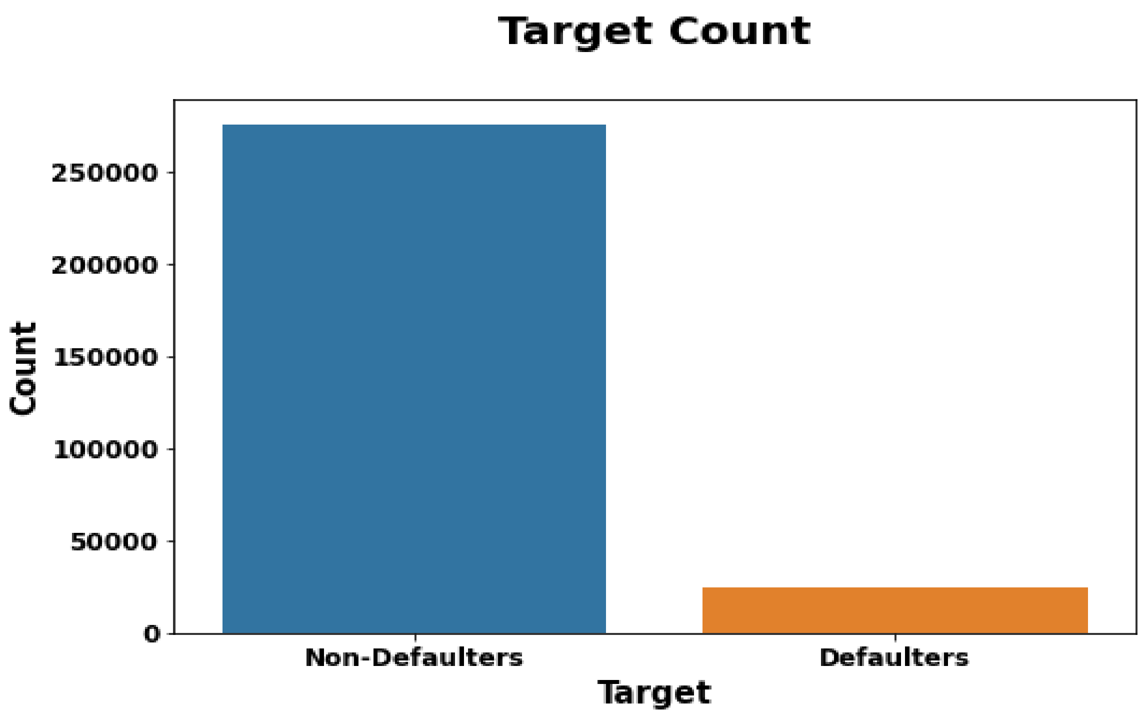


Result:

Although the resulted plots still contains outliers even after we deleted some records, it is still continuous in nature. Hence these can be treated as normal values.

2.2 Checking Data Imbalance

TARGET VARIABLE

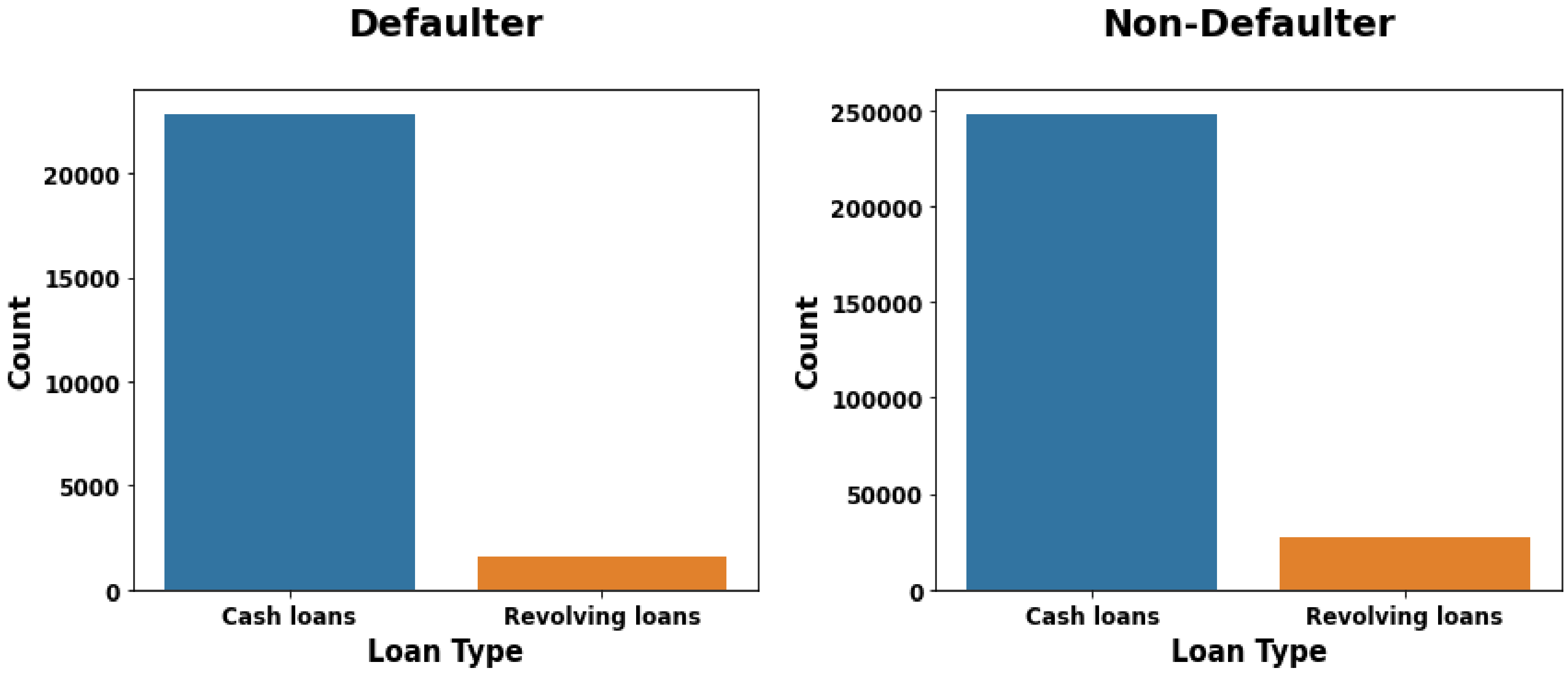


Observation:

We can clearly see that there is high data imbalance between **Defaulters** and **Non-Defaulters**. We have 91.82% clients marked as **Non-defaulters** and rest 8.18% marked as **Defaulters**.

2.3 Univariate and Segmented Univariate Analysis

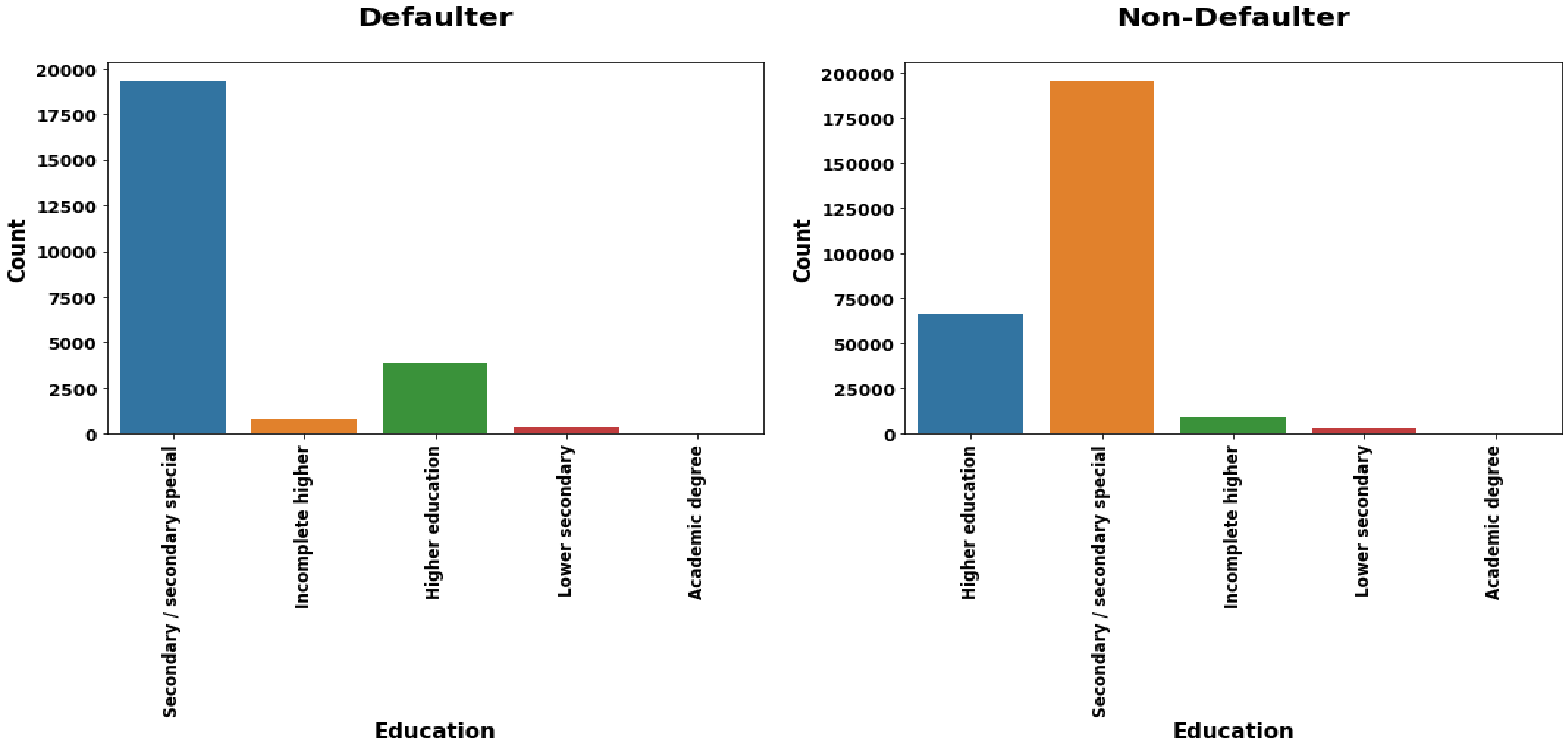
Type of Loan



Observation:

We can clearly see that most of the clients prefers applying for cash loans than revolving loans in both Defaulters and Non-Defaulters cases.

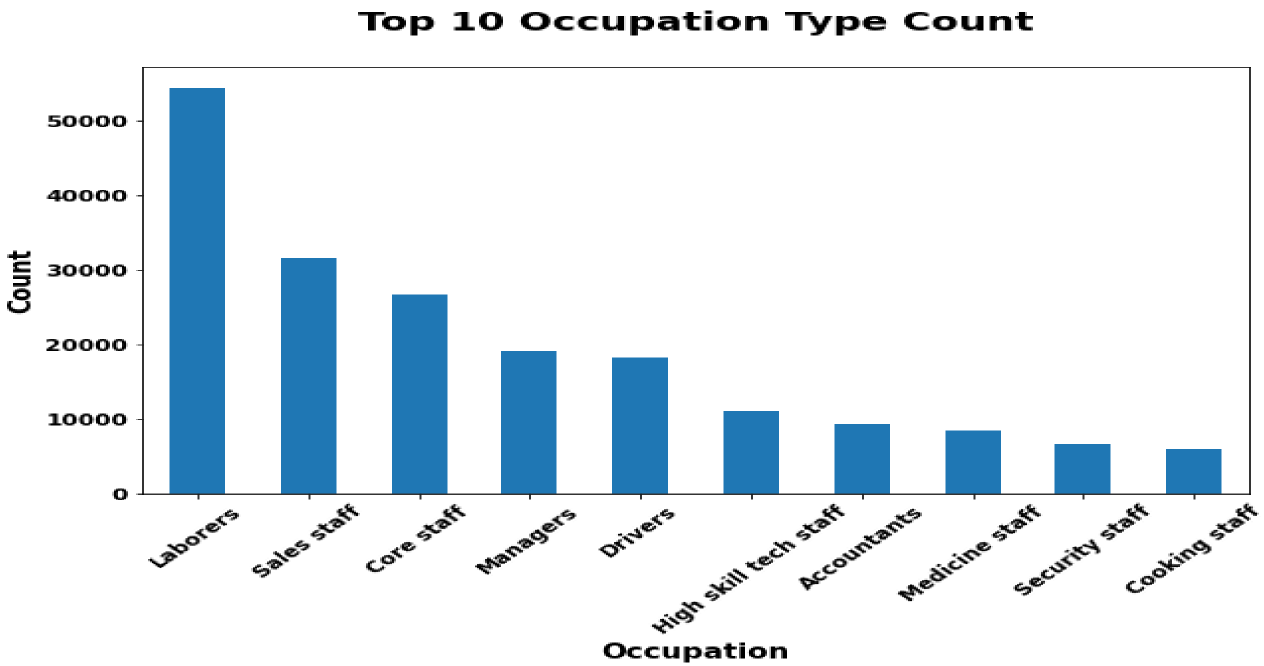
Client Educational Background



Observation:

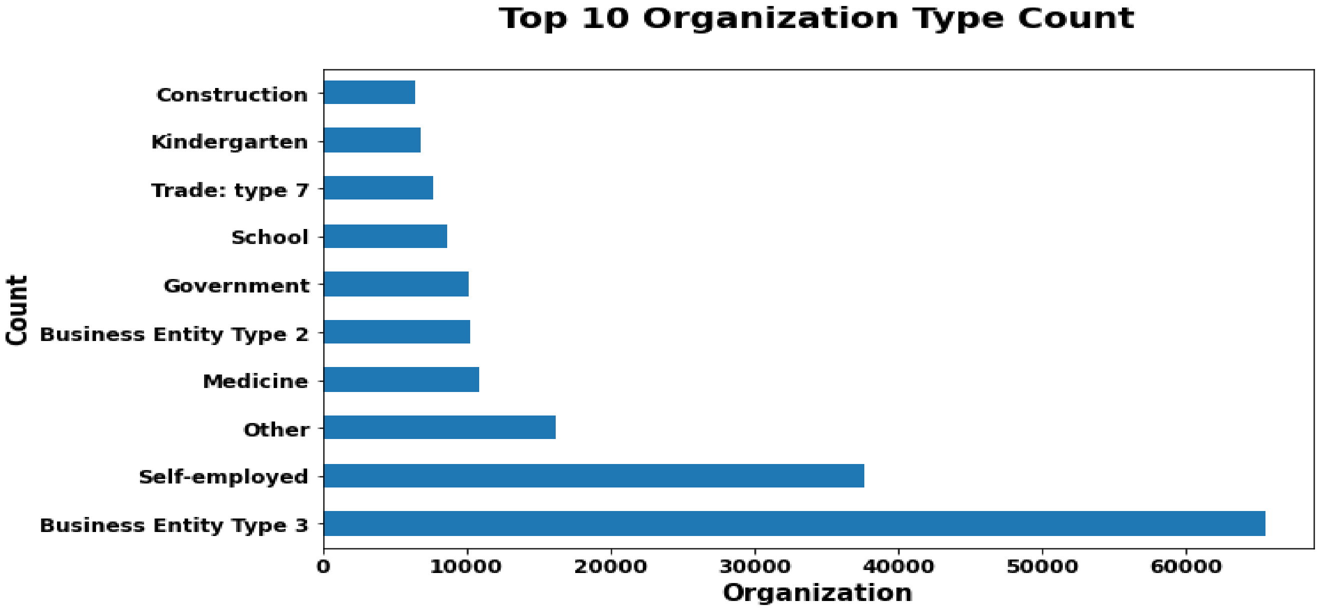
We can see that most of the clients who cleared bank loans have pursued secondary education followed by higher education.

Client Occupation and Organization Background



Observation:

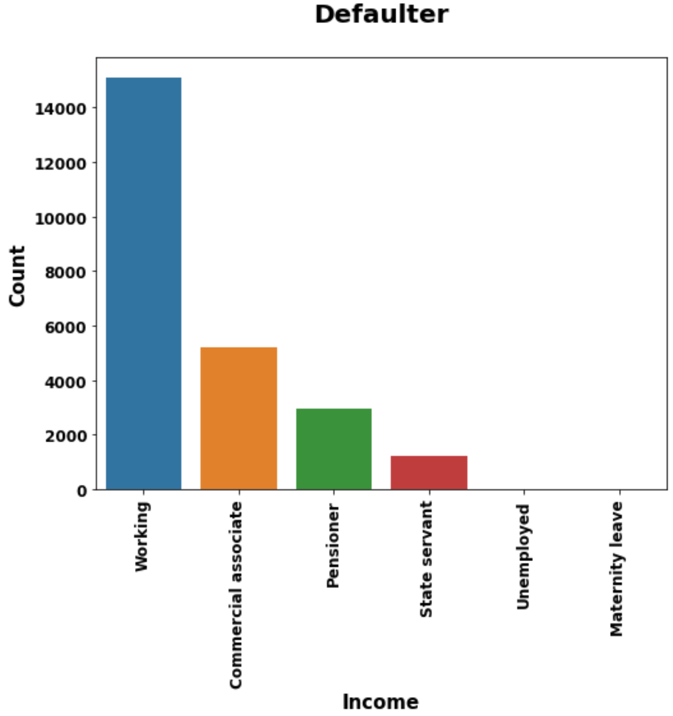
We have most of the clients with Occupation as **Laborers** from top 10 list who shown more interest for loan followed by **Sales staff**, **Core staff** and **Managers**.

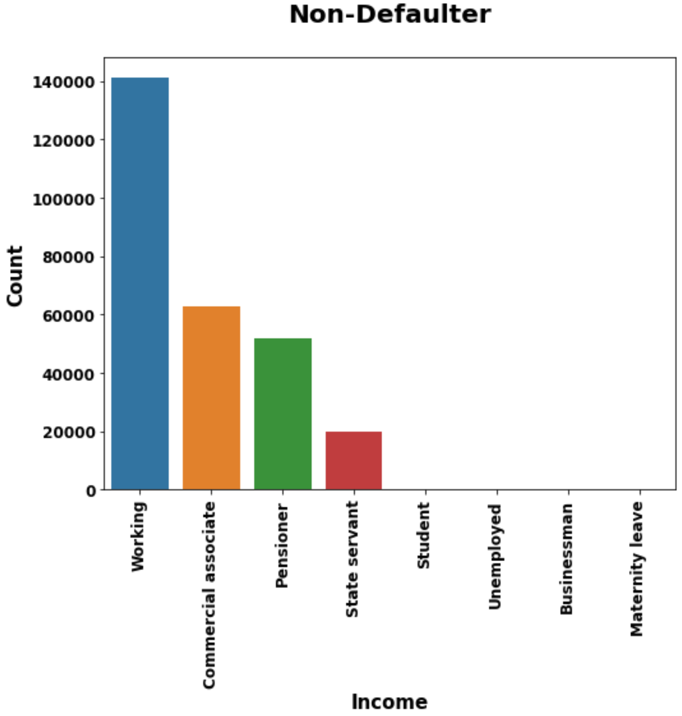


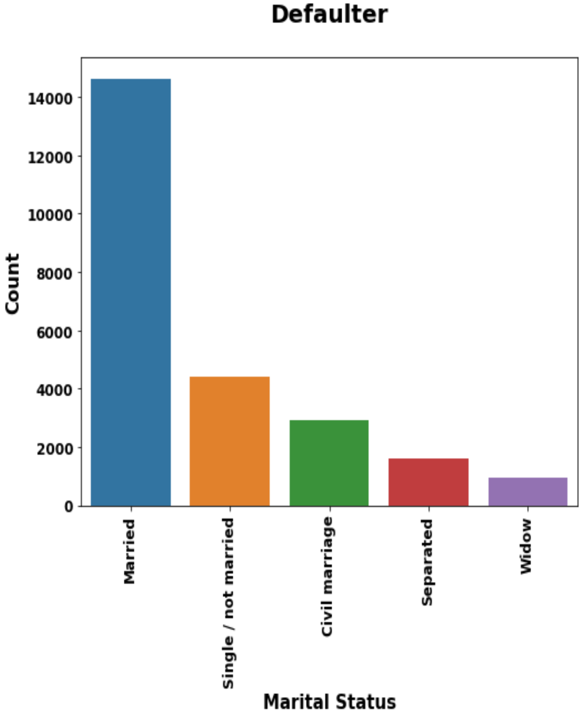
Observation:

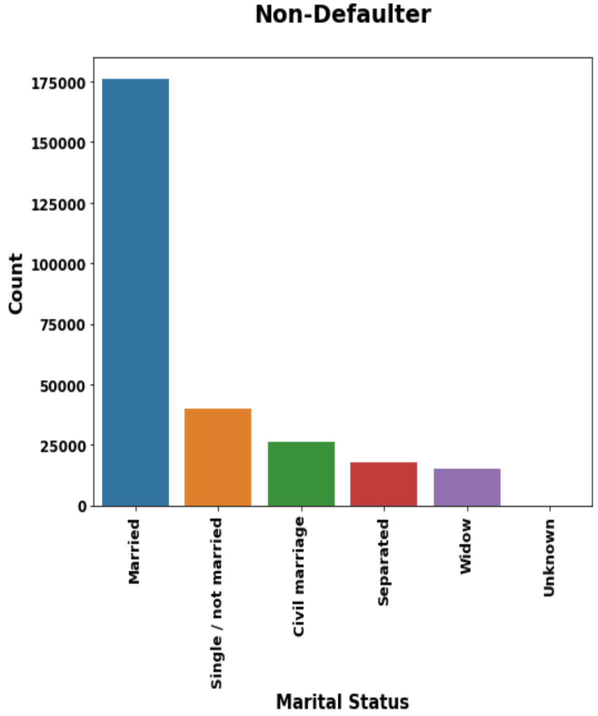
We have most of the clients belongs to **Business Entity Type 3** class from top 10 list who shown more interest for loan followed by **Self-employed**.

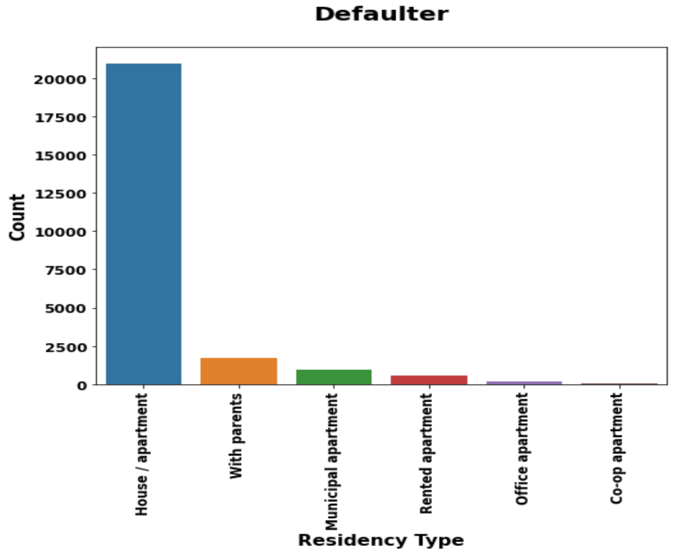
Client Status Information

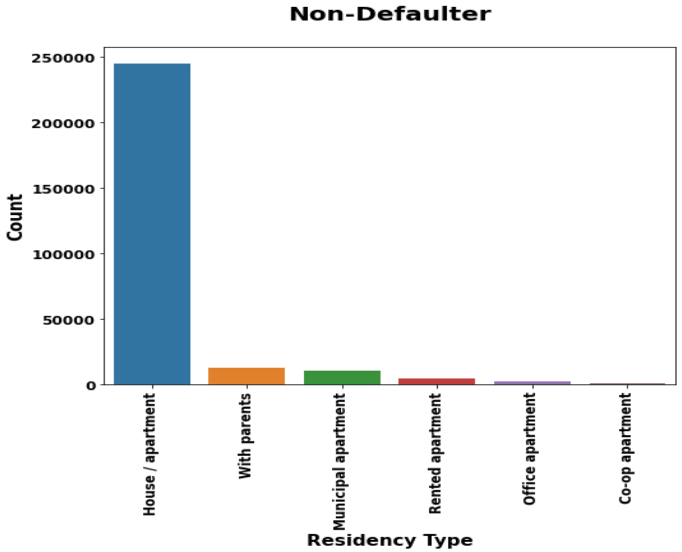










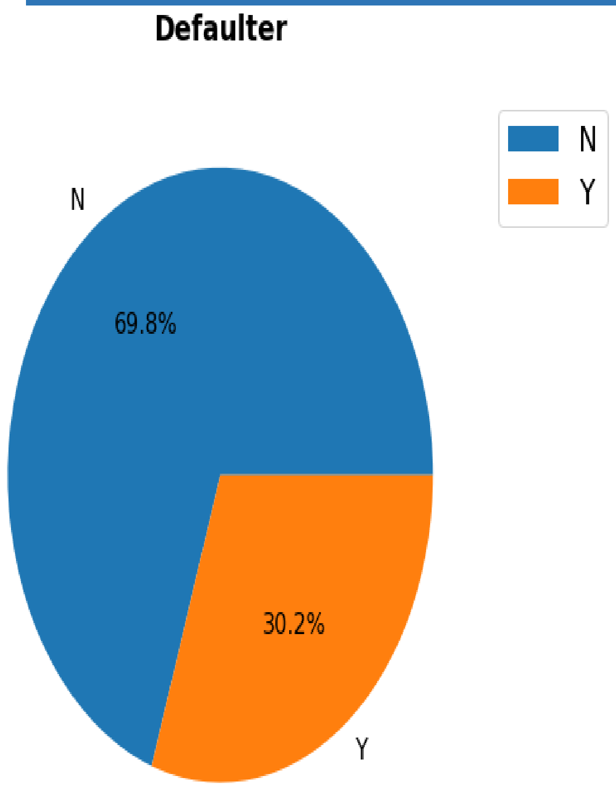


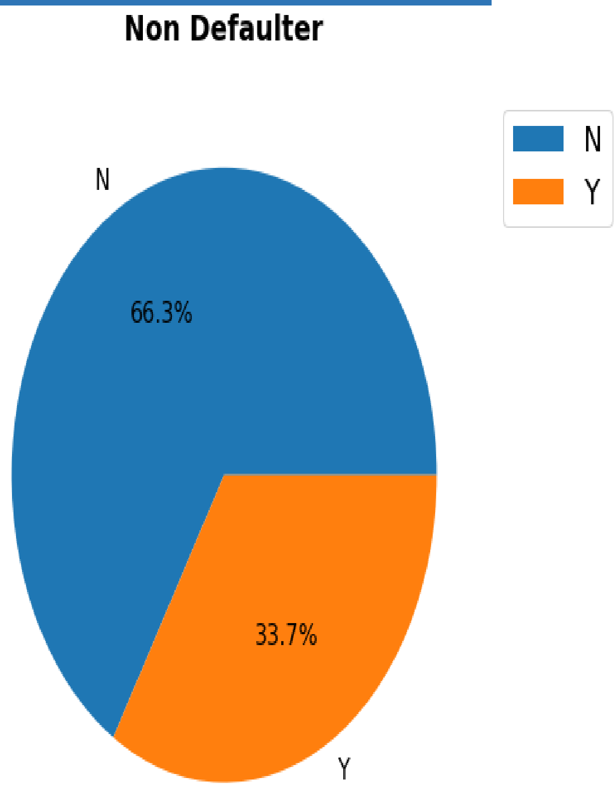
Observation:

- Most of clients (both defaulters and non-defaulters) belongs to **"Working** group.
- Next best group/class is **Commercial associate** followed by **Pensioner** and **State servant**.
- Most of the clients (Defaulters and Non-Defaulters) are **Married**.
- Next best group is **Single** followed by **Civil marriage**.
- Most of the clients are **residing in house/appartment** and have **high data imbalance** compared with other types for Defaulters and Non-Defaulters.
- Few clients comes under the group **staying with parents** followed by **Municipal apartment, Rented apartment**. Very few clients residing in **office apartment and co-op apartment**.

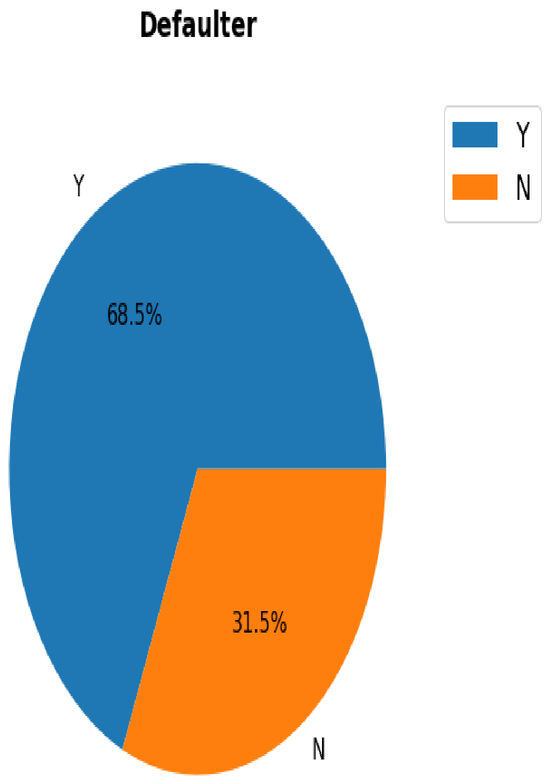
Client Asset Details

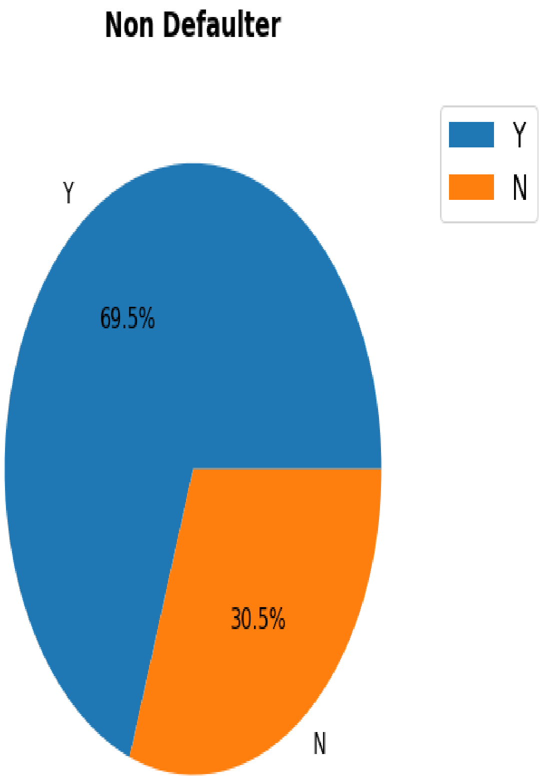
Client Own Car





Client Own House



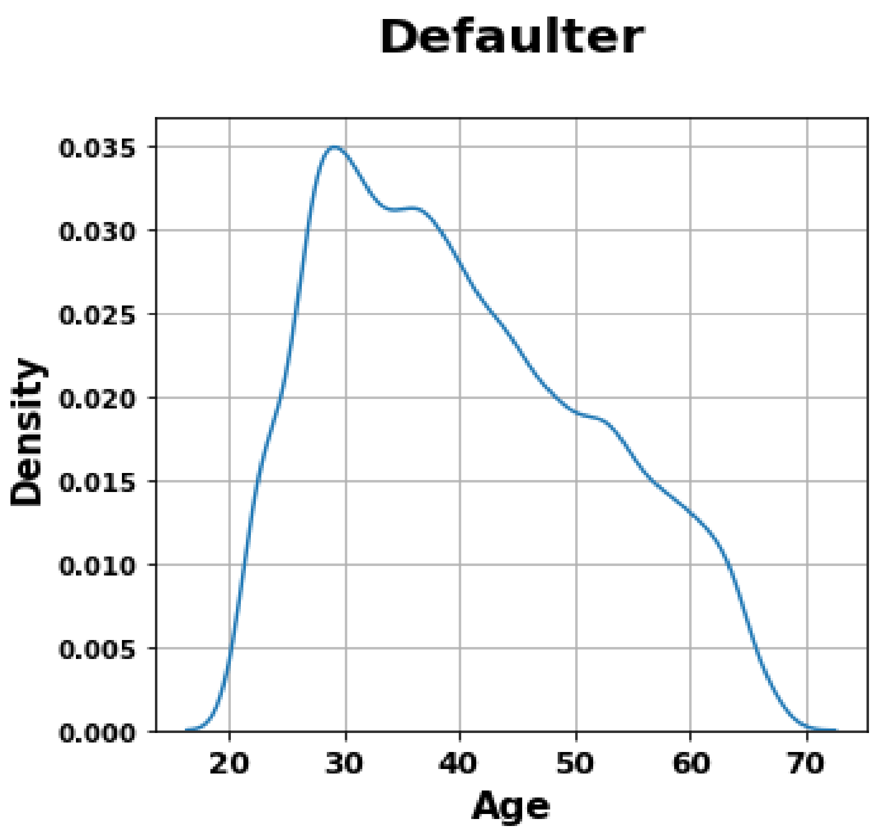


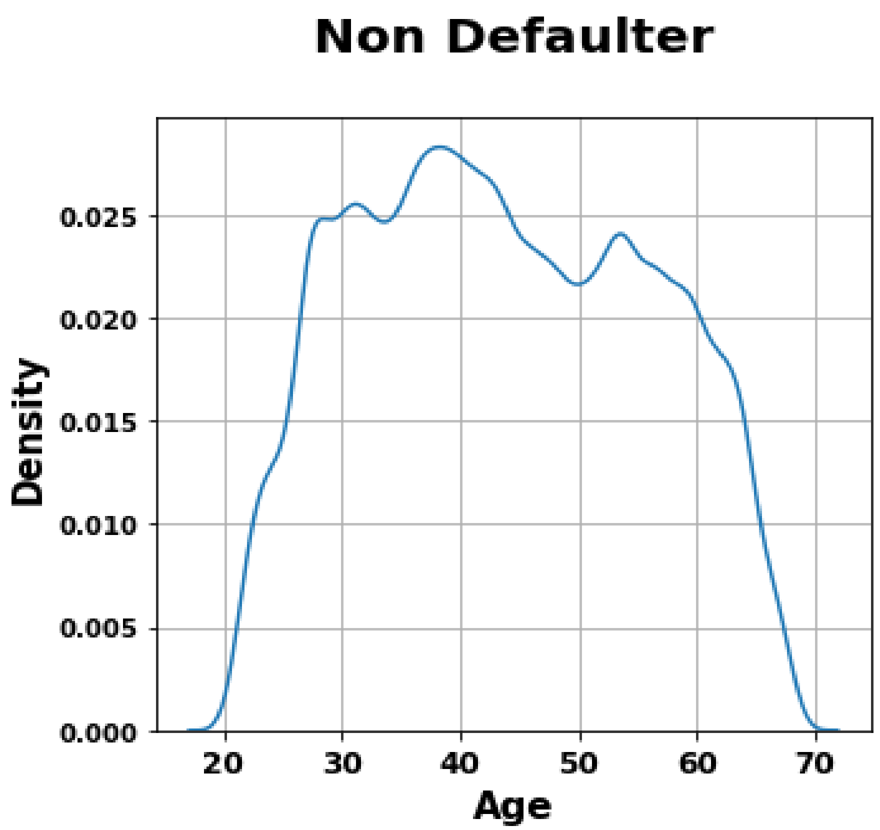
Observation:

From the above plots we can say that:

- We have around **69.8%** clients from the total number of defaulters who **doesn't own car**.
- We have around **31.5%** clients from the list of number of defaulter who **doesn't own house**.
- In both the assets, defaulters have less ratio than non-defaulters.

Client Age Group



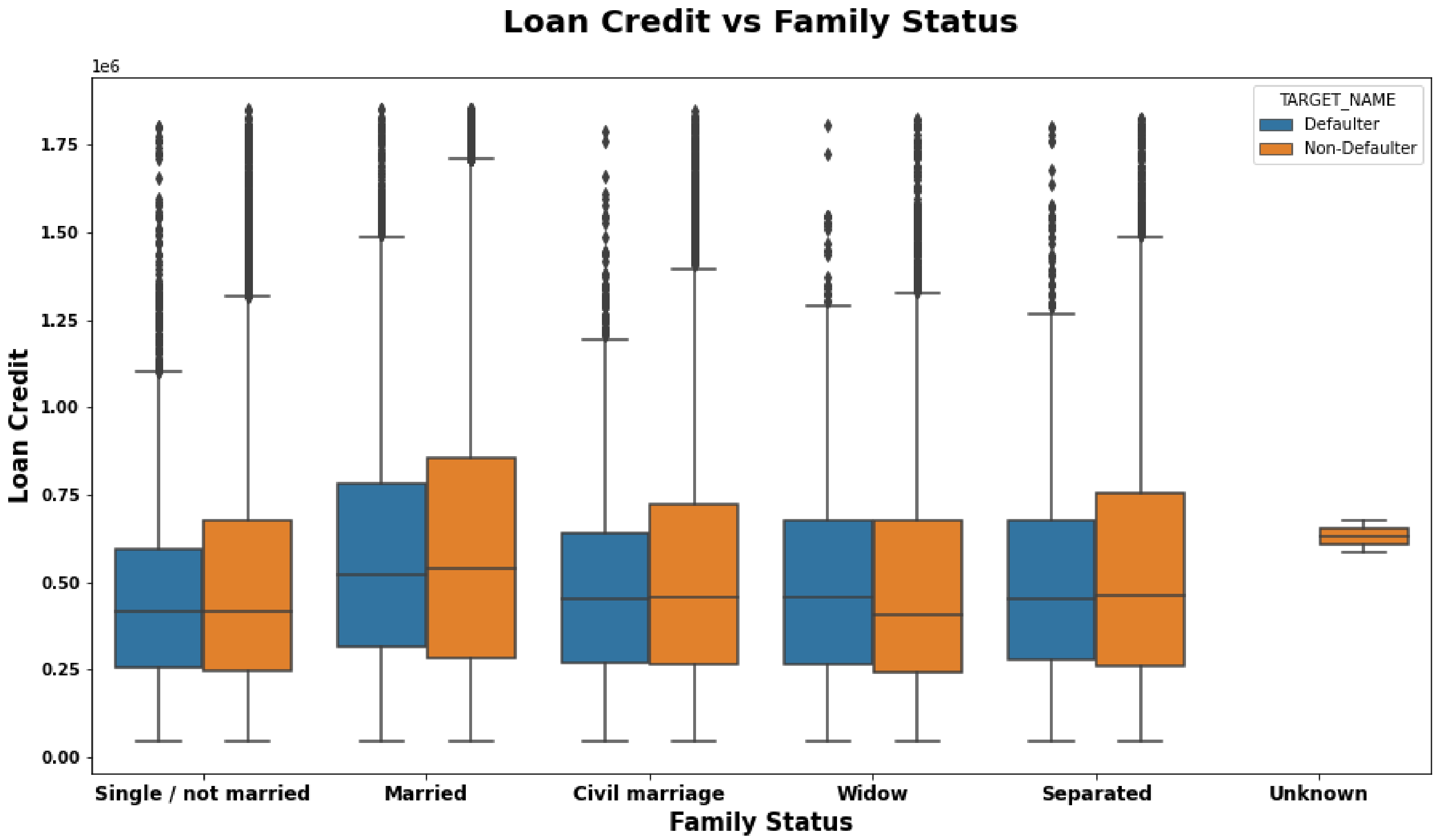


Observation:

- In defaulter case,
  - Most of the clients belong to age group between 25 and 40.
  - Number of clients gradually decreases when age increases.
- In Non-defaulter case,
  - Most of the clients belong to age group between 25 and 55.
  - Very few clients are present from the age 60.

2.4 Bivariate and Multivariate Analysis

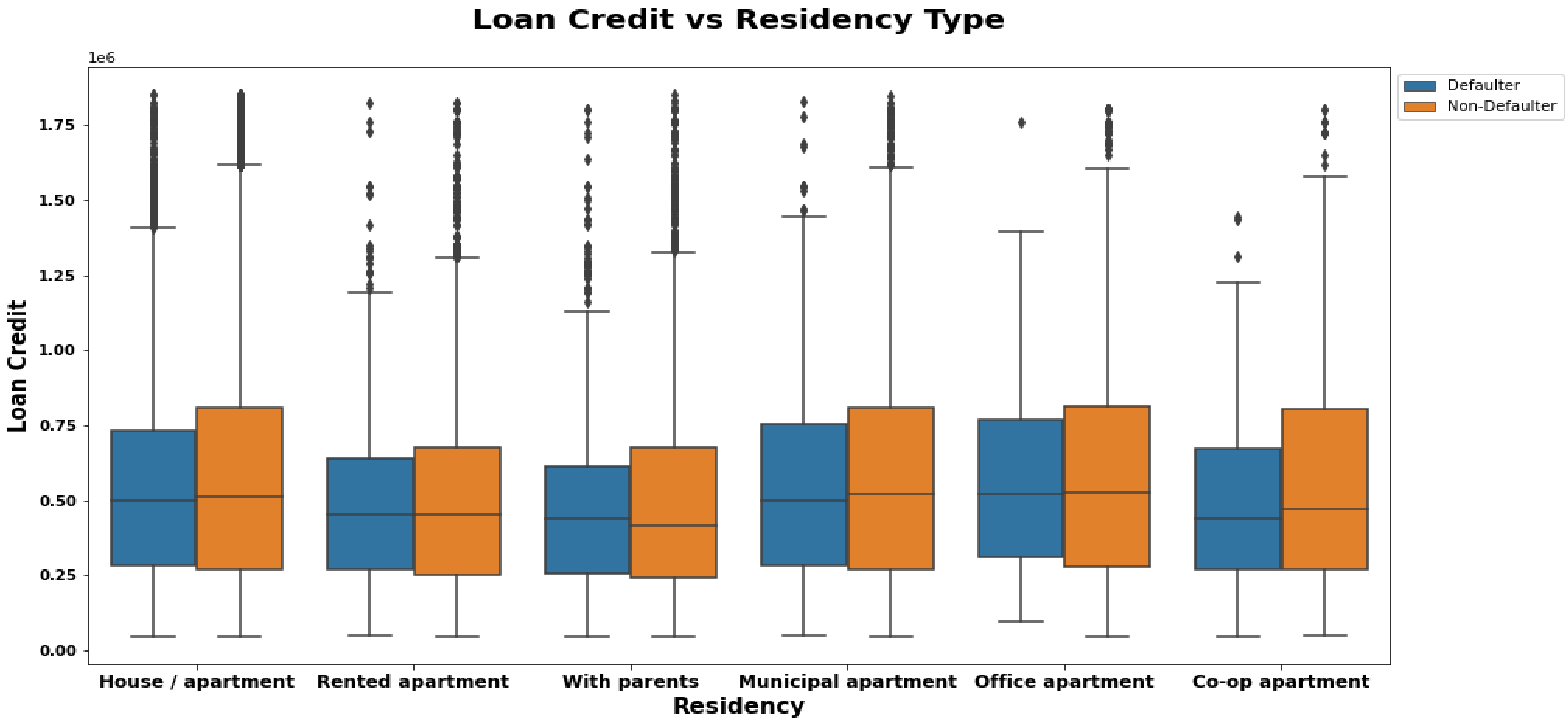
Loan Credit Sanctioned for each Family Status type w.r.t Target Variable



Observation:

- **Highest and Average** loan amount credited for **Married** status for both defaulters and non-defaulters are having **higher** in number than other status types (excluding Unknown).
- Average loan amount credited for the client who belongs to **Widow** status for defaulters is **greater** than non-defaulters. Bank has to consider this factor while providing loan to this type.

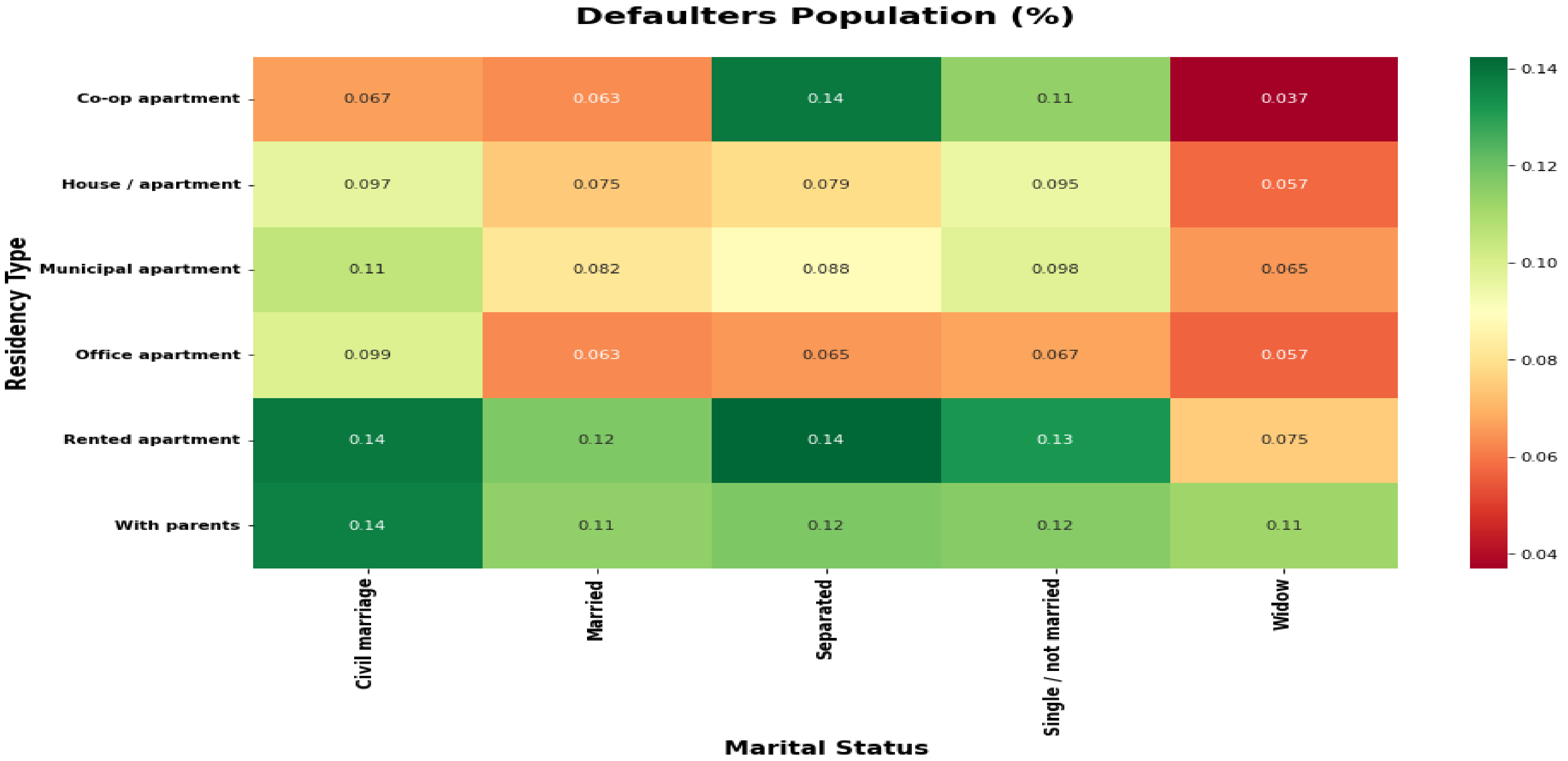
Loan Credit Sanctioned for each type of Clients Residency w.r.t. Target Variable



Observation:

- Average loan amount sanctioned is similar for **House/Apartment, Municipal apartment and Office apartment** types.
- For both defaulters and non-defaulters, highest loan credit provided for the client who resides in **House/apartment**.
- Defaulters who provided with high loan credit and lives in house/apartment are the **most** in population and continuous in nature who haven't cleared loan on time compared with other residency types. Bank has to consider this while providing loan to this type.

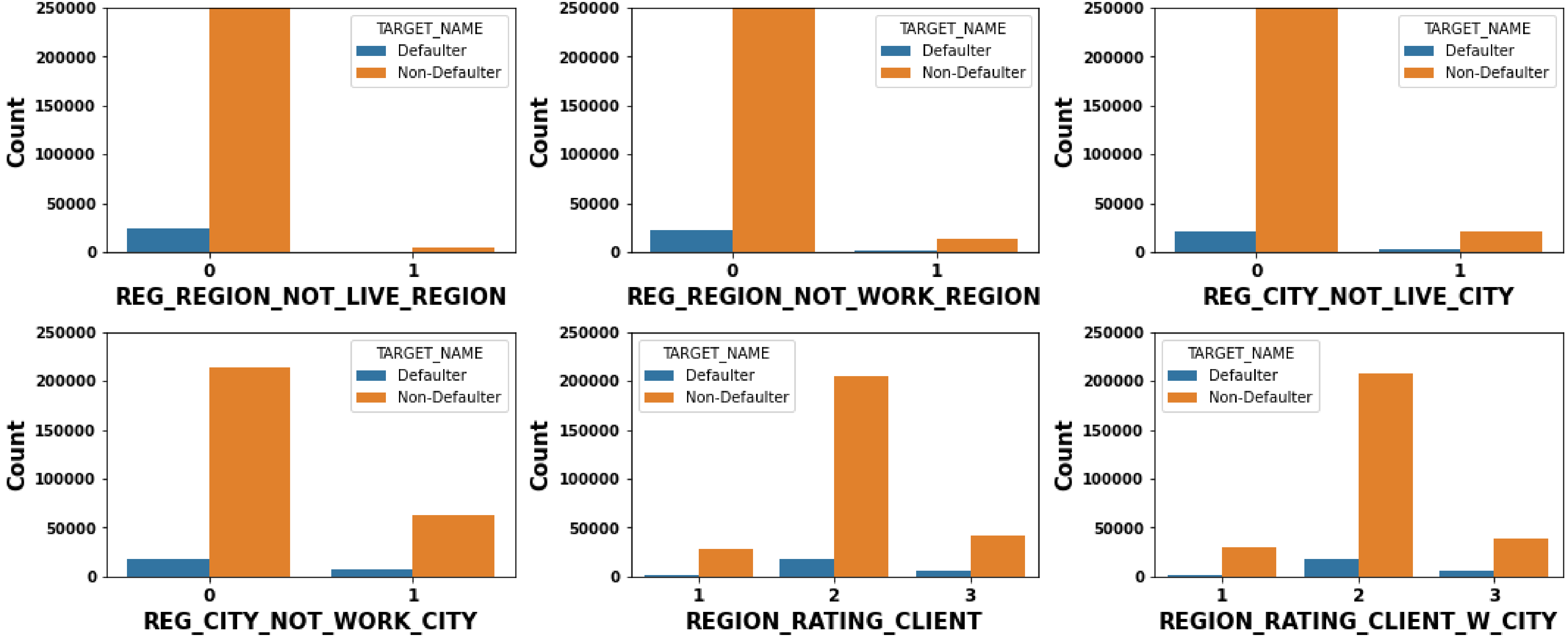
Check Defaulters Population in each Family Status and Residency Type



Observation:

- From the above plot we can clearly say that a greater number of defaulters are staying either in **Rented apartment or with parents especially** whose marital status is **civil marriage, separated or single/not married**
- We also have decent number of defaulters with **Married or Widow status** who resides in **rented apartment or with parents**.
- We also have high number of defaulters under co-op apartment residency type with marital status as **separated or single/not married**.
- We have less amount of defaulters population for other residency types.

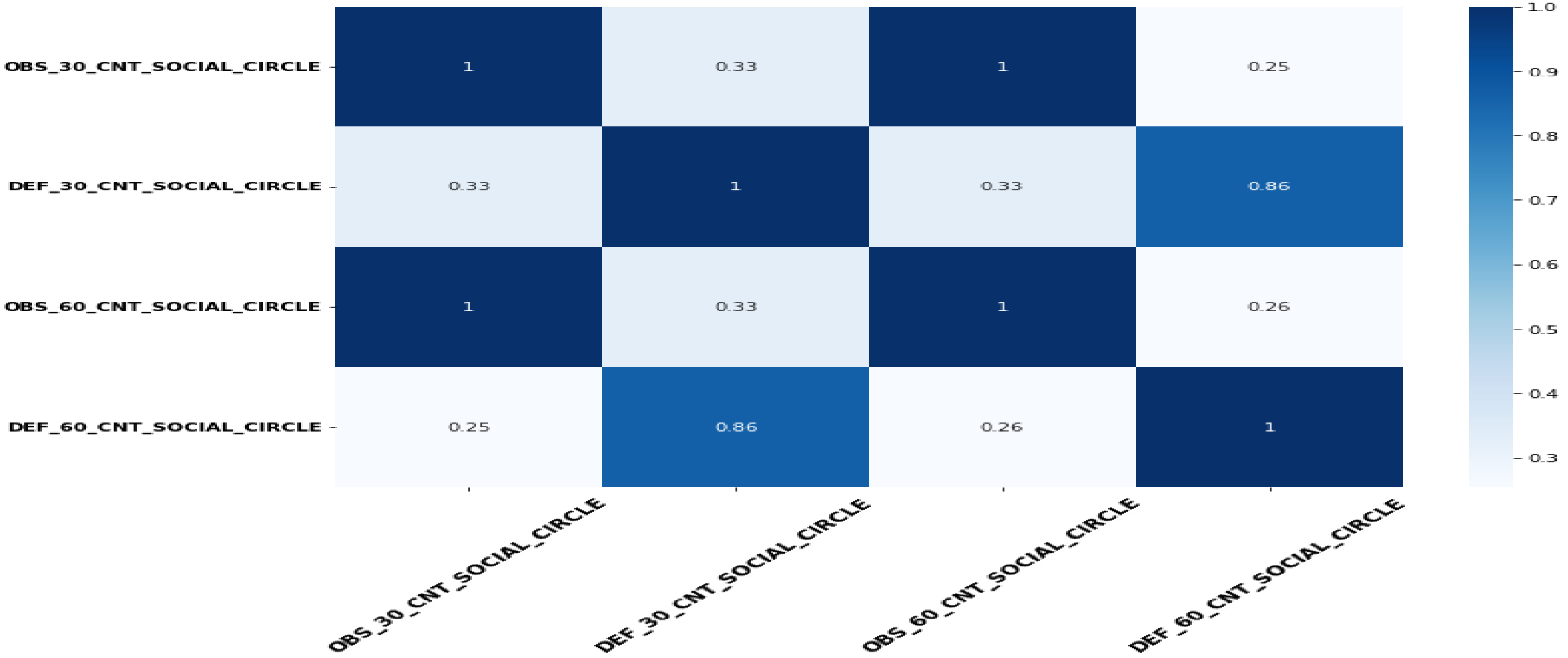
Region Related Information



Observation:

- Defaulter rate is highest when REG_REGION_NOT_WORK_REGION=0, REG_REGION_NOT_LIVE_REGION=0, REG_CITY_NOT_LIVE_CITY=0, REG_CITY_NOT_WORK_CITY=0 i.e. **permanent address and working address is same**.
- More number of Applicants/clients have **Region rating of 2** (both region and city levels).

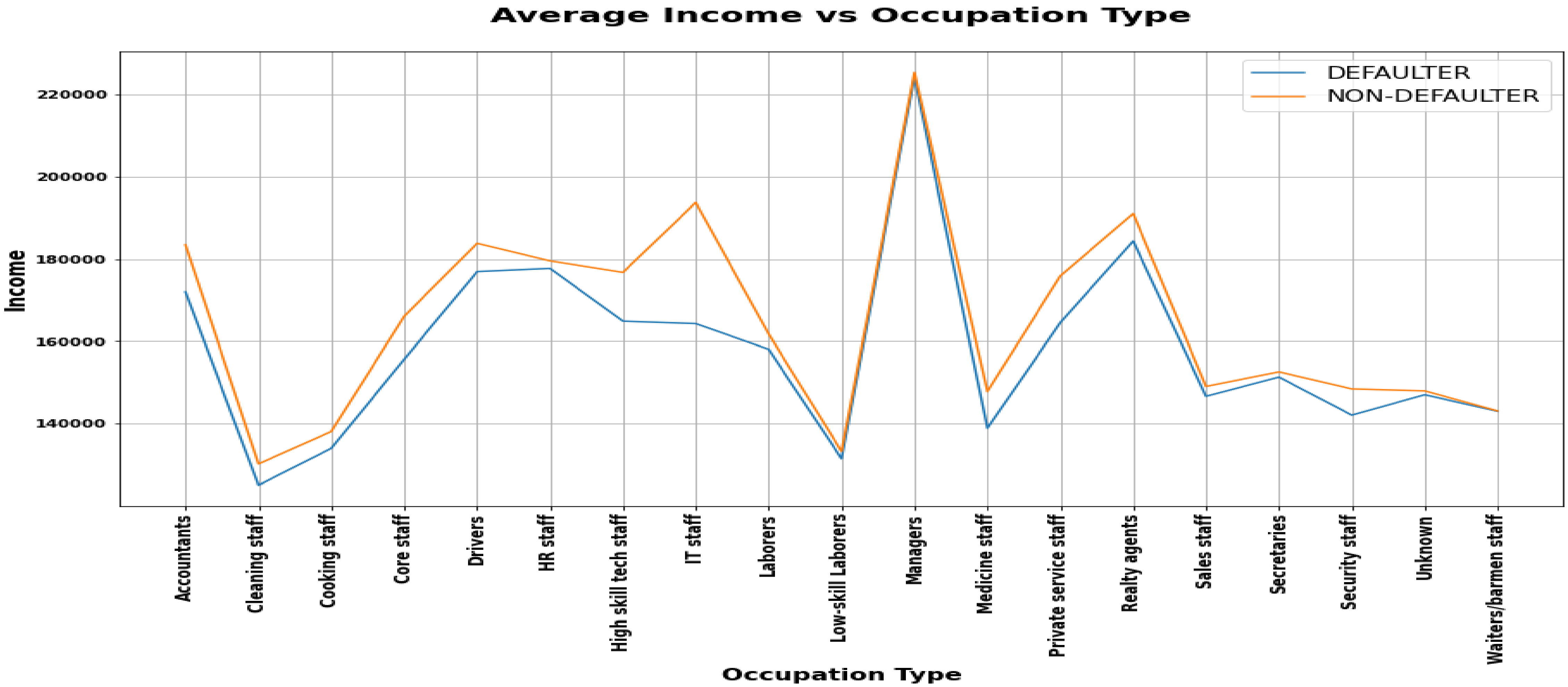
Client Social Circle



Observation:

- DEF_30_CNT_SOCIAL_CIRCLE and DEF_60_CNT_SOCIAL_CIRCLE are highly correlated i.e., if the number of observations of client's social surroundings defaulted on 30 DPD increases then the number of observations of client's social surroundings defaulted on 60 DPD also increases.
- OBS_30_CNT_SOCIAL_CIRCLE and OBS_60_CNT_SOCIAL_CIRCLE are identical columns and we can have any one of the feature in consideration and other can be ignored.

Average Income Salary for each Occupation Type w.r.t Target variable



Observation:

- We can see that **Managers** have highest average income salary compared with other occupation types for both Defaulters and Non-Defaulters.
- Next best highest average income salary present for **IT staff** followed by **Realty agents, Accountants and Drivers**.
- Defaulters having less average income for all the occupation types compared with Non-Defaulters.

2.5 Top 10 Correlation for Client with Payment Difficulties and Other Cases

Top 10 Correlation Variables for Defaulters (Client with Payment Difficulties)

1. (OBS_30_CNT_SOCIAL_CIRCLE, OBS_60_CNT_SOCIAL_CIRCLE)
2. (AMT_CREDIT, AMT_GOODS_PRICE)
3. (REGION_RATING_CLIENT_W_CITY, REGION_RATING_CLIENT)
4. (CNT_FAM_MEMBERS, CNT_CHILDREN)
5. (DEF_30_CNT_SOCIAL_CIRCLE, DEF_60_CNT_SOCIAL_CIRCLE)
6. (LIVE_REGION_NOT_WORK_REGION, REG_REGION_NOT_WORK_REGION)
7. (REG_CITY_NOT_WORK_CITY, LIVE_CITY_NOT_WORK_CITY)
8. (AMT_CREDIT, AMT_ANNUITY)
9. (AMT_ANNUITY, AMT_GOODS_PRICE)
10. (FLAG_DOCUMENT_6, DAYS_EMPLOYED)

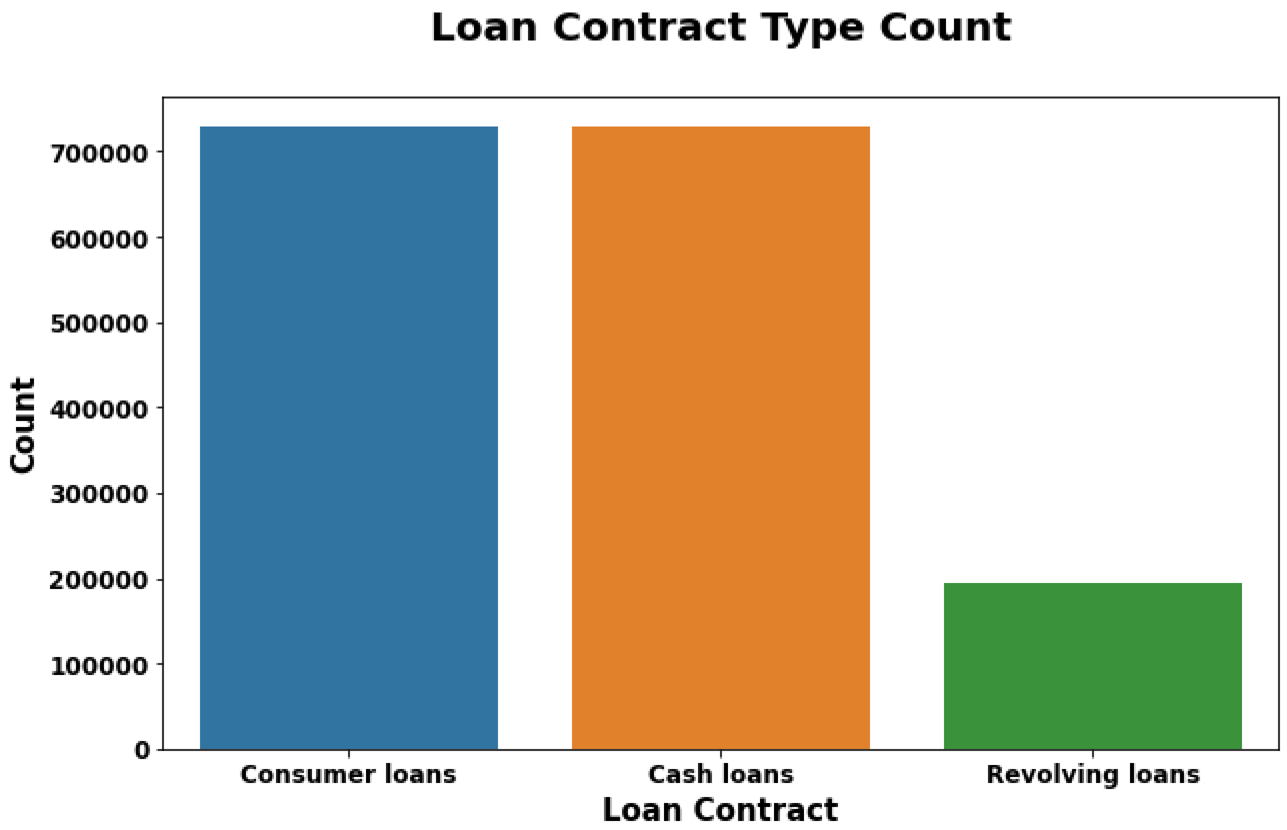
Top 10 Correlation Variables for Non-Defaulters (Client without any Payment Difficulties)

1. (OBS_30_CNT_SOCIAL_CIRCLE, OBS_60_CNT_SOCIAL_CIRCLE)
2. (AMT_GOODS_PRICE, AMT_CREDIT)
3. (REGION_RATING_CLIENT_W_CITY, REGION_RATING_CLIENT)
4. (CNT_CHILDREN, CNT_FAM_MEMBERS)
5. (REG_REGION_NOT_WORK_REGION, LIVE_REGION_NOT_WORK_REGION)
6. (DEF_60_CNT_SOCIAL_CIRCLE, DEF_30_CNT_SOCIAL_CIRCLE)
7. (LIVE_CITY_NOT_WORK_CITY, REG_CITY_NOT_WORK_CITY)
8. (AMT_GOODS_PRICE, AMT_ANNUITY)
9. (AMT_ANNUITY, AMT_CREDIT)
10. (DAYS_EMPLOYED, DAYS_BIRTH)

3. Previous Loan Application Data Analysis

3.1 Univariate Analysis

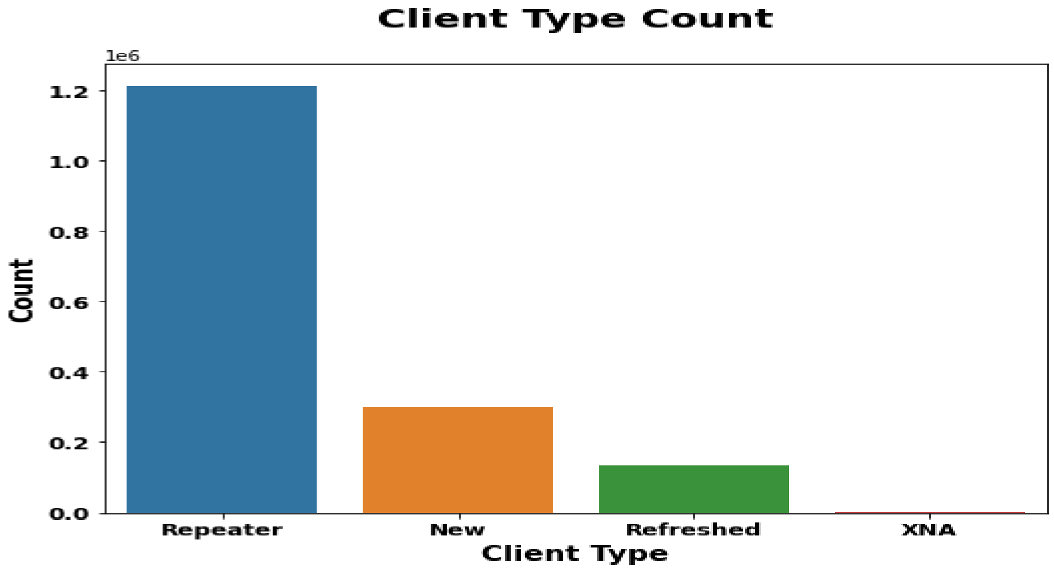
Loan Contract Type In Previous Application

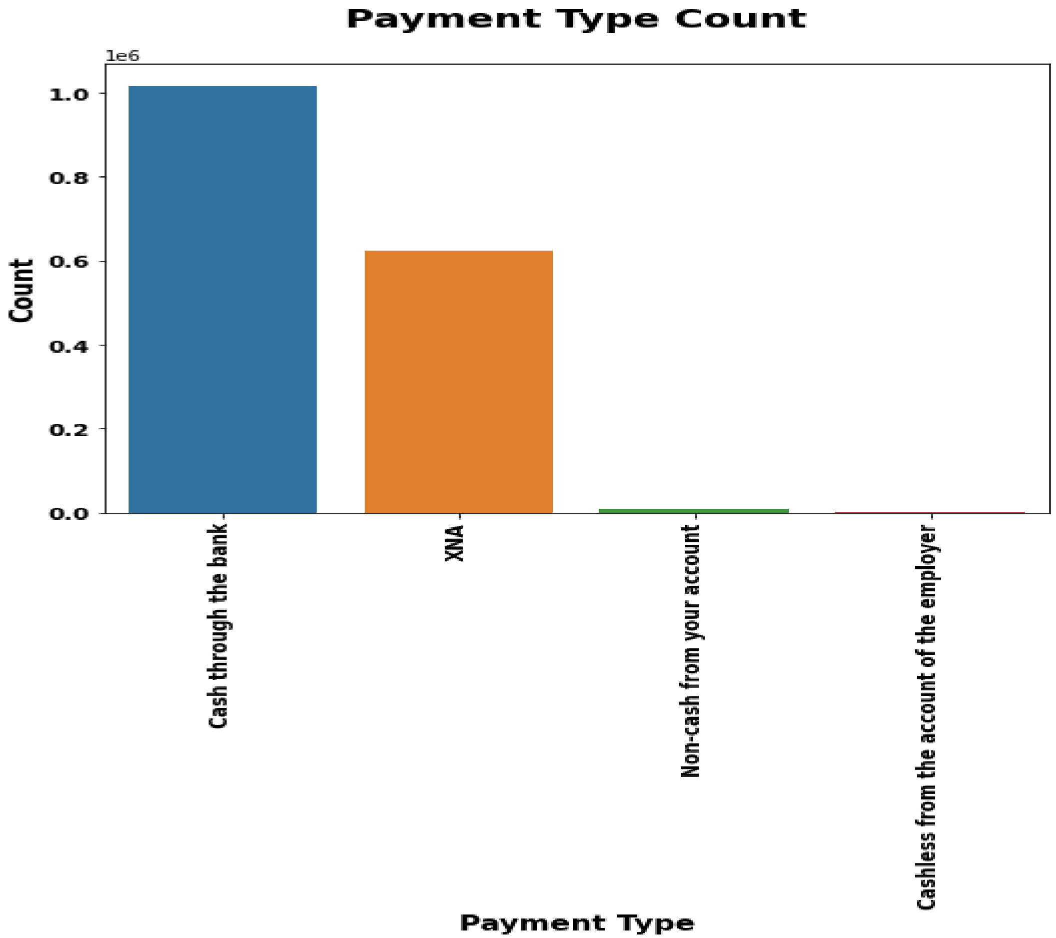


Observation:

- We can say that majority of the clients applied either **Consumer loans or Cash loans type**. Applicants counts for both of these clients are almost similar.
- We have a smaller number of applicants/clients who applied for **Revolving loans**.

Client Previous Application Type and Mode of Payment Preferred

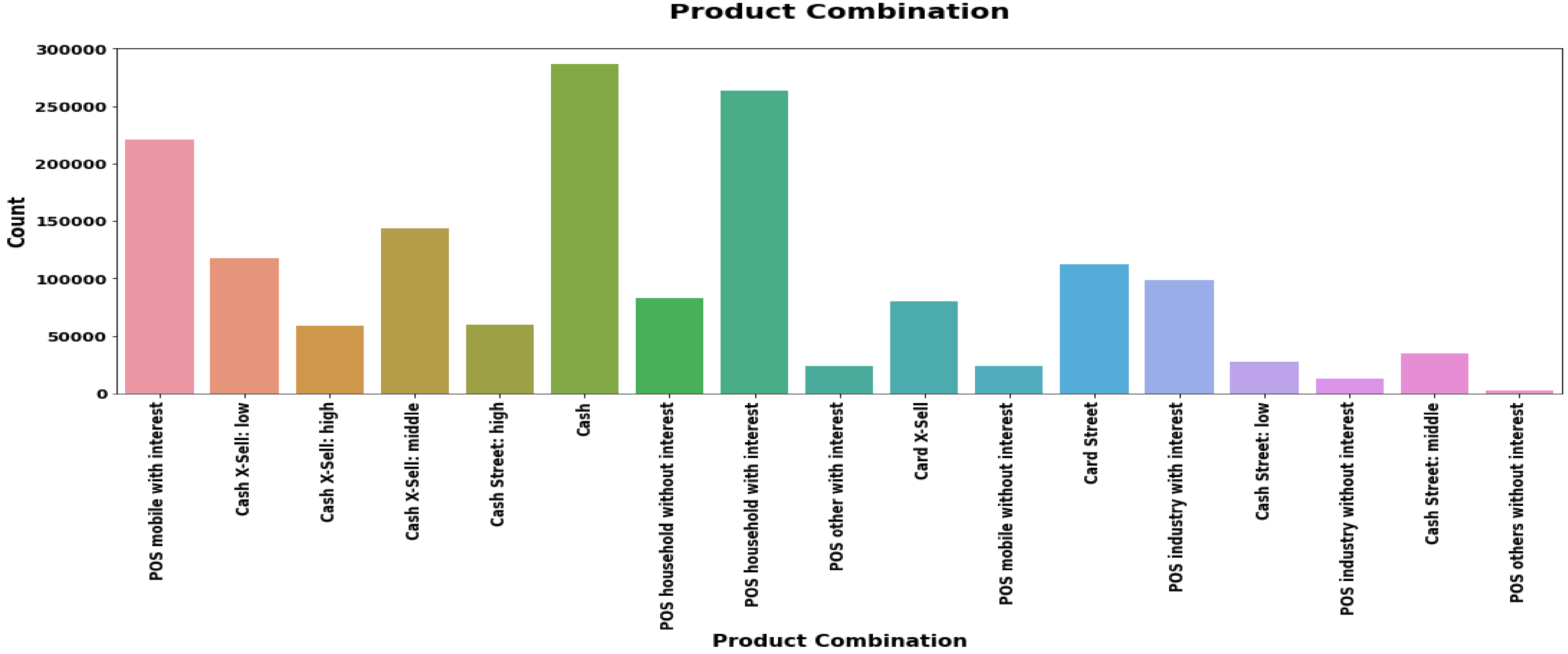




Observation:

- Most of the clients are **repeater**.
- **'Cash through the bank'** is the most frequently used payment method.

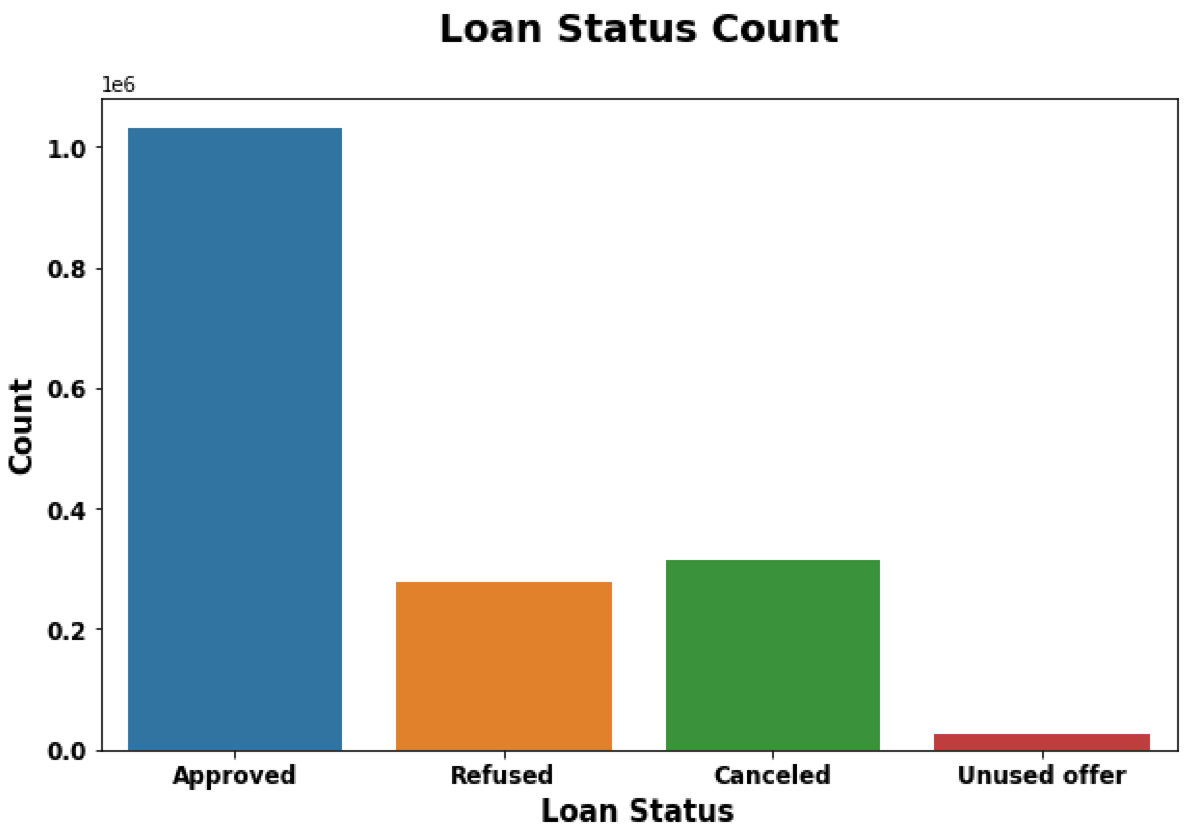
Product Combination Type For Which Loan Applied

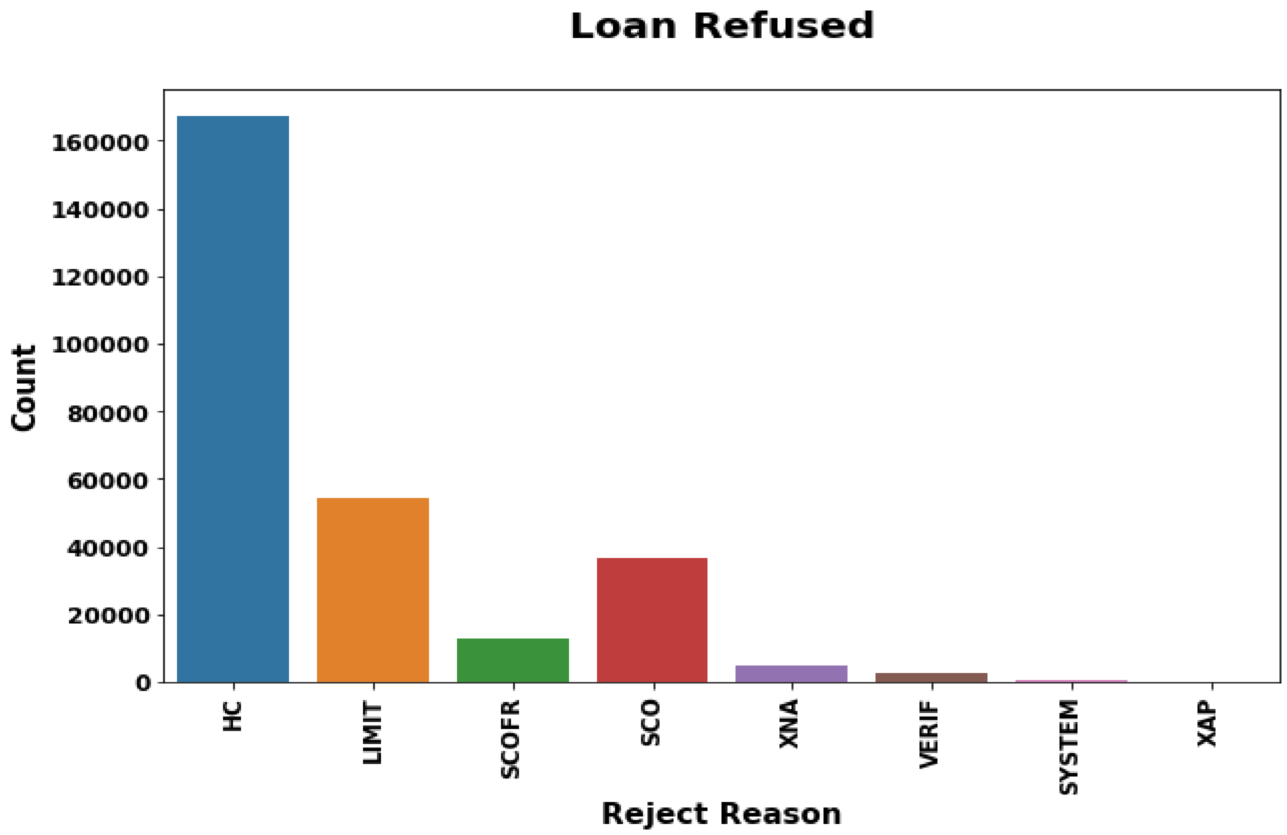


Observation:

- Highest number of clients applied for **cash** loan followed by **POS household with interest and POS mobile with interest** products.

Loan Application Status and Reason For Rejection





Observation:

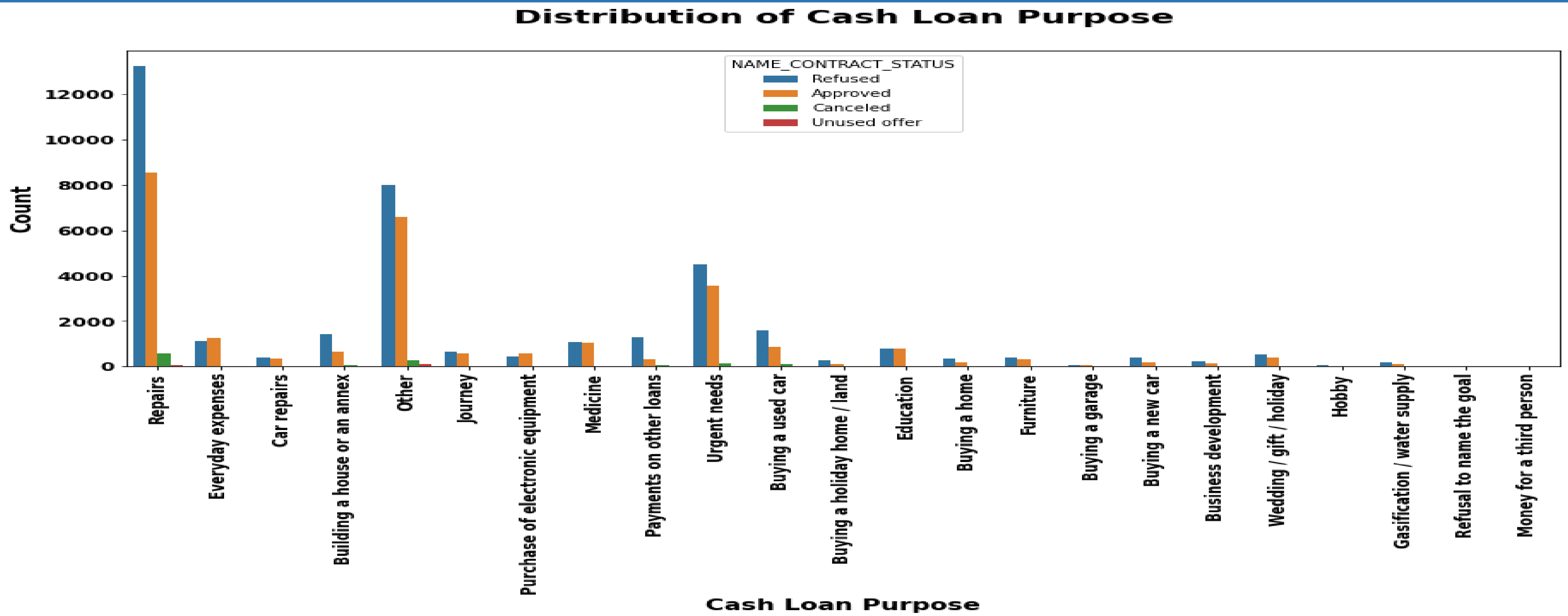
- Majority of the previous loan applications got **approved**.
- Some of the other applications are in **cancelled or refused** status.
- Very less applications are in **Unused offer** status.

Observation:

- **'HC', 'LIMIT' and 'SCO'** (descending order) are the most common reasons for loan refused/rejected.

3.2 Bivariate and Multivariate Analysis

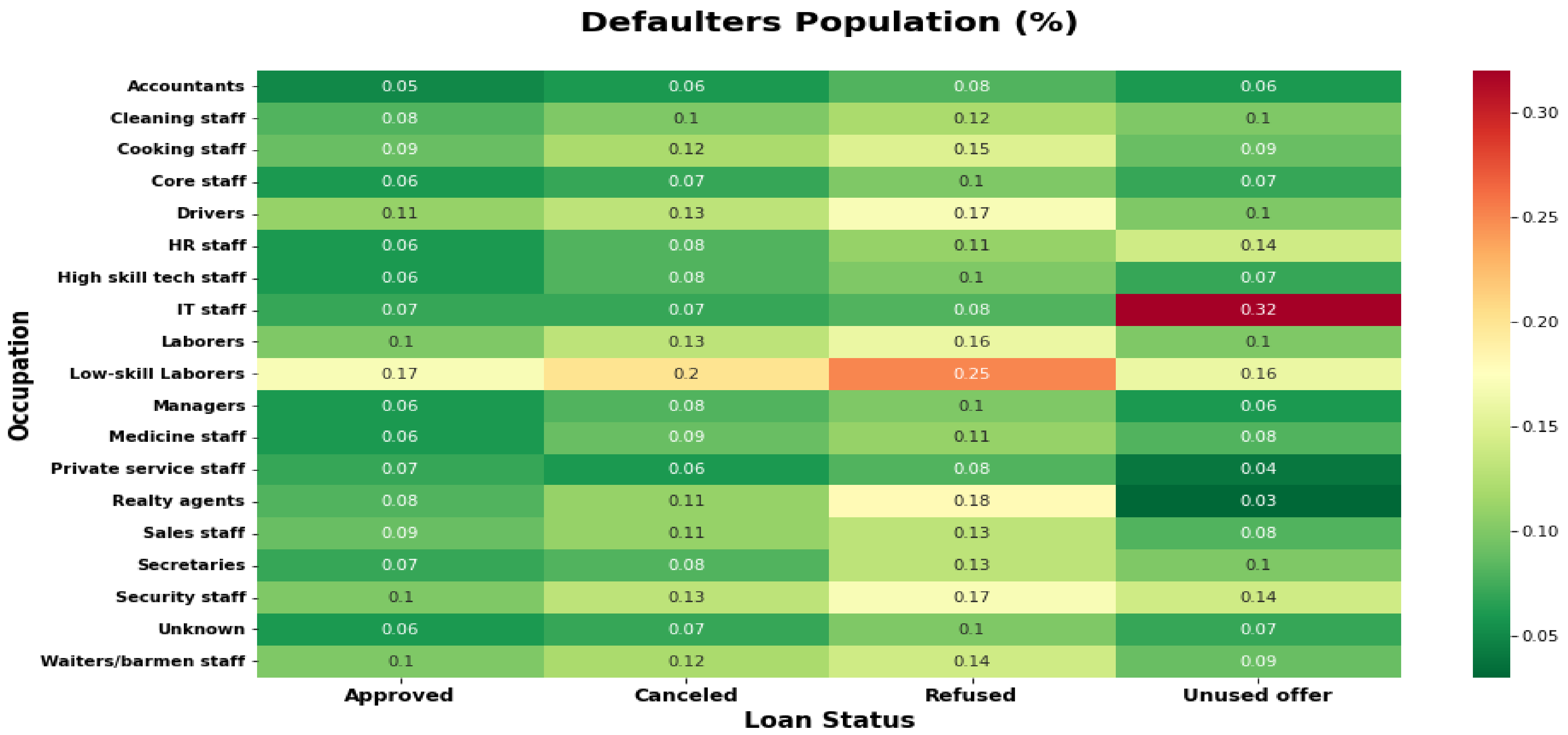
Loan Contract Status For Each Cash Loan Type



Observation:

- Maximum number of the loans are rejected which have loan purpose as **Repairs**.
- **Payments of other loans** has much higher rejection than approval.
- **Education loan** has almost equal rates of approvals and rejections.

Check Defaulters From Previous Application - Occupation vs Loan Status



Observation:

- More number of defaulters loan applications got **rejected** whose occupation is **Low skilled Laborers**.
- Less number of defaulters loan applications got **rejected** whose occupations are **Private service staff, IT staff, Accountant and Manager**.

4. Conclusion

Bank/Finance Company must consider the following factors before providing loan to the clients:

1. From the experience/previous data, a client whose occupation is of Manager, IT staff, Accountant and Private Service Staff are less defaulters and able to clear the loans without any issue.
2. High loan repayment success rate from the clients whose age is between 25 and 50. This makes sense as most of them are working professionals with good income salary.
3. Provide loan to the client who own assets like own car or house will be a good consideration and high chance of loan repayment from them.
4. Though 'Repair' Purpose have higher chance of loan repayment, but they also have highest chance of being a defaulter. So, need to be more cautious in paying the loan for 'Repair' Purpose.
5. Beware of providing loan to low skilled laborers as they had high percentage of defaulter's rate in the previous applications.
6. Client's Income type is the most important driving factor to decide loan sanction. It gives us a risk estimate of loan repayment. Do the background verification such as client's working profession, living region, current family status etc., before providing the loan.