EXPLORATORY DATA ANALYSIS OF LOAN APPLICATION

By Shreyas Singhal

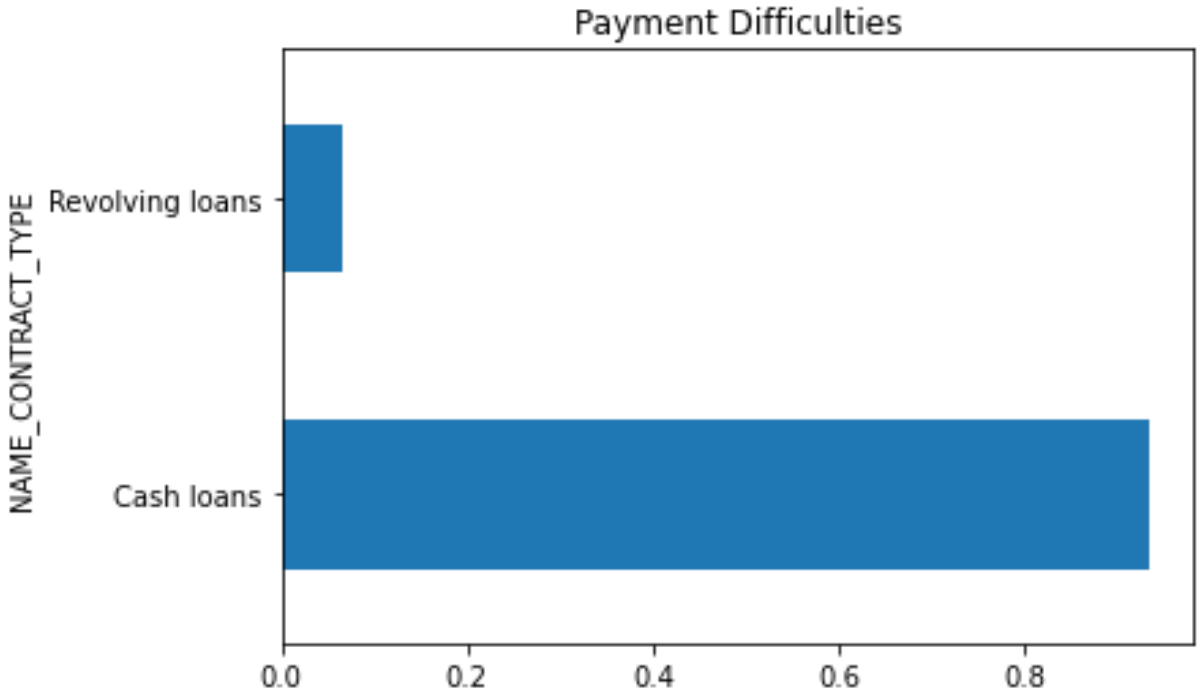
PROBLEM STATEMENT

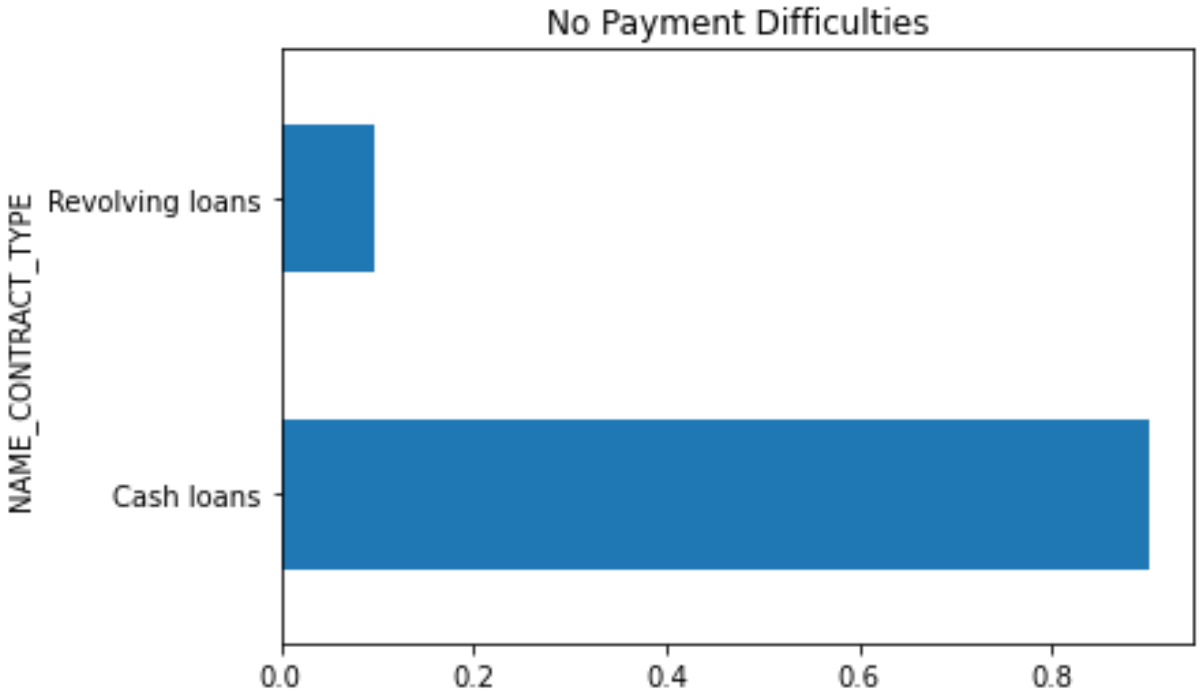
- In the traditional banking methods the loan approval process was extremely slow and risky.
- Before credit history was also difficult to identify.
- If a customers credit history was good then bank has no problem in giving loans to him, but if the credit history is not present and then too bank provide them loans then customer could use the same as their advantage by becoming a defaulter.
- But at the same time if the customer is capable of paying the loan and bank doesn’t give them loans this would result in a loss to bank.
- As the technology is growing we need to find an Intelligent Business Solution to overcome this problem.
- Exploratory Data Analysis is one of the Intelligent Solutions.
- By considering the Past Data, applying the Past Data into Present Data we can predict the future.
- Future is whether giving loan to a particular customer will result in loss or profit to bank.
- Whether that specific customer will repay the loan or not in future.

Overall Approach

- First we will analyze the application data and get the insights from the data.
- First step is Data Cleaning.
- We will first delete the Null Values (<30%) in the data and accordingly replace the pending missing values with Mean, Median or Mode depending on the situation.
- Checking the data types and check if the data types are correct or not.
- Deleting columns which are unnecessary and not required for the analysis.
- Second Step is Analysis.
- With the help of graphs we will analyze the data and get as much insights possible.
- Second we will analyze the previous application data and get the insights from the data.
- Then again we will follow the steps mentioned from point no. 2.
- We will combine the Previous Application Data and Current Application Data.
- Accordingly we will delete the unnecessary columns and use the necessary columns to get meaningful insight from the data.
- Using different graphs we will analyze the data.

CONTRACT TYPE



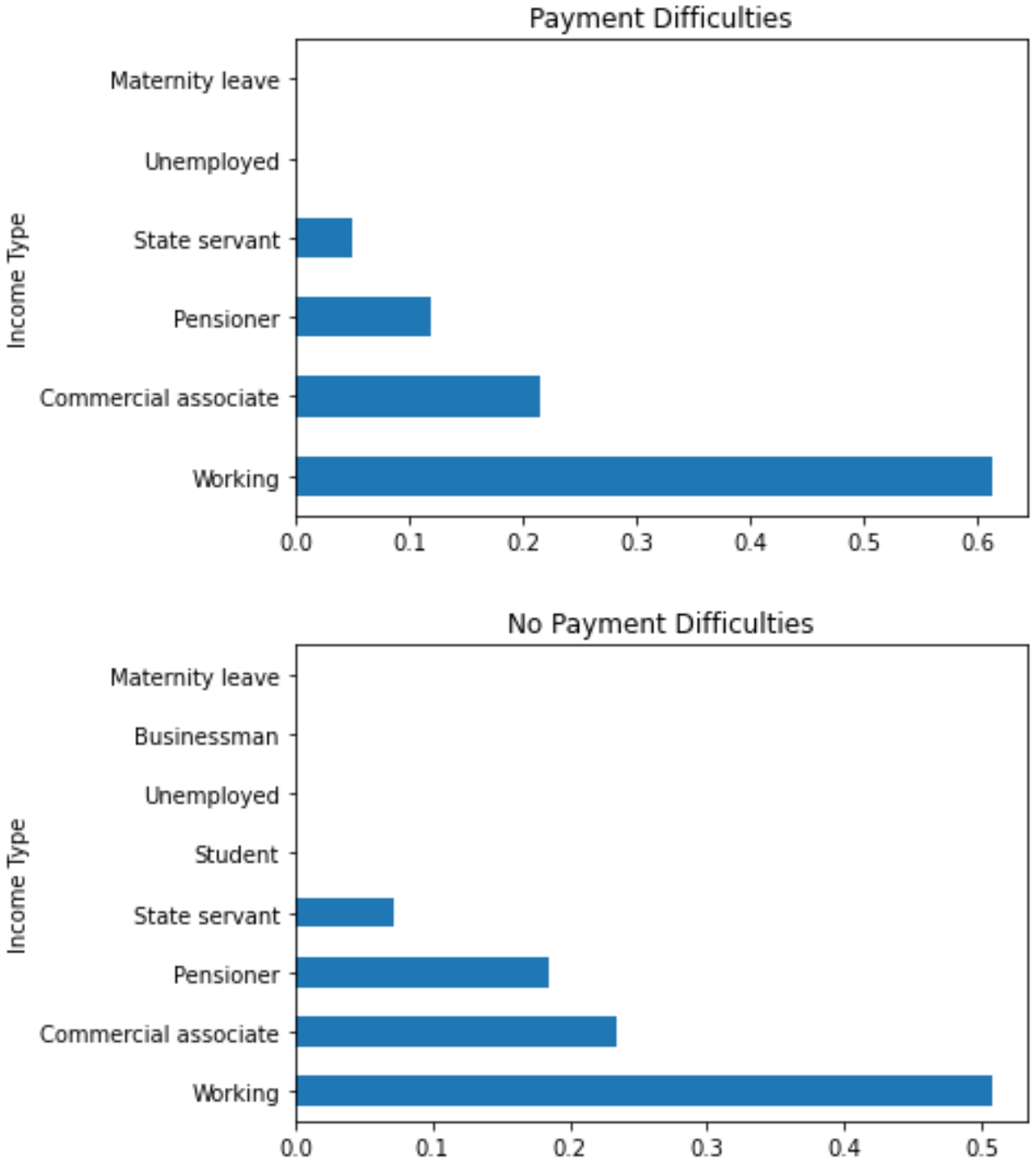


After observing both the graphs we conclude the following:

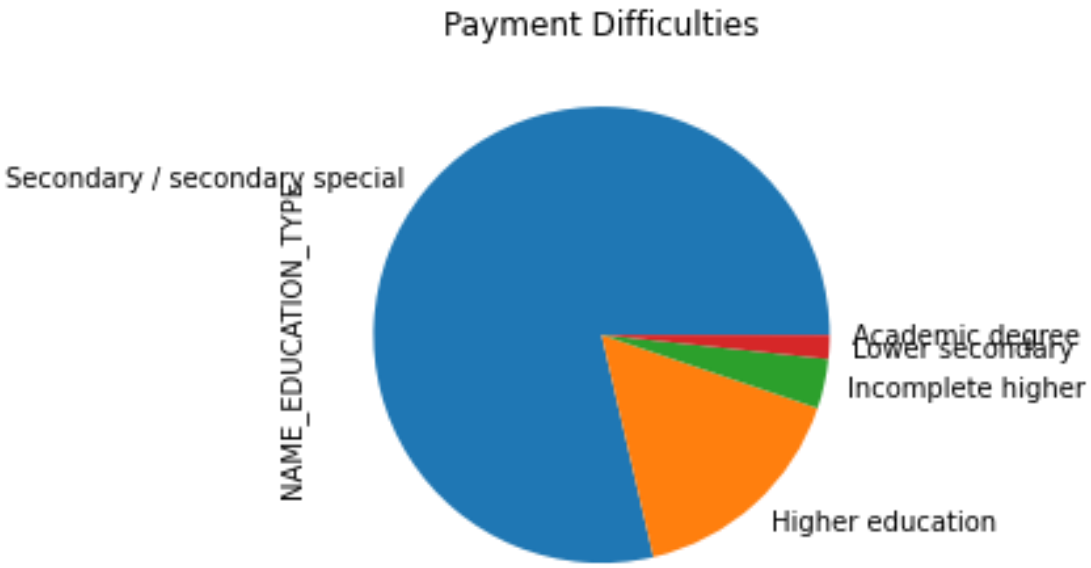
- We can conclude that the bank face more payment difficulties in Cash Loan Type Compared to Revolving Loans Type.
- Also we can see that Cash Loan Type customers tend to pay on time Compared to Revolving Loans Type.
- At the same time bank has maximum contracts with Cash Loan Type Customers.

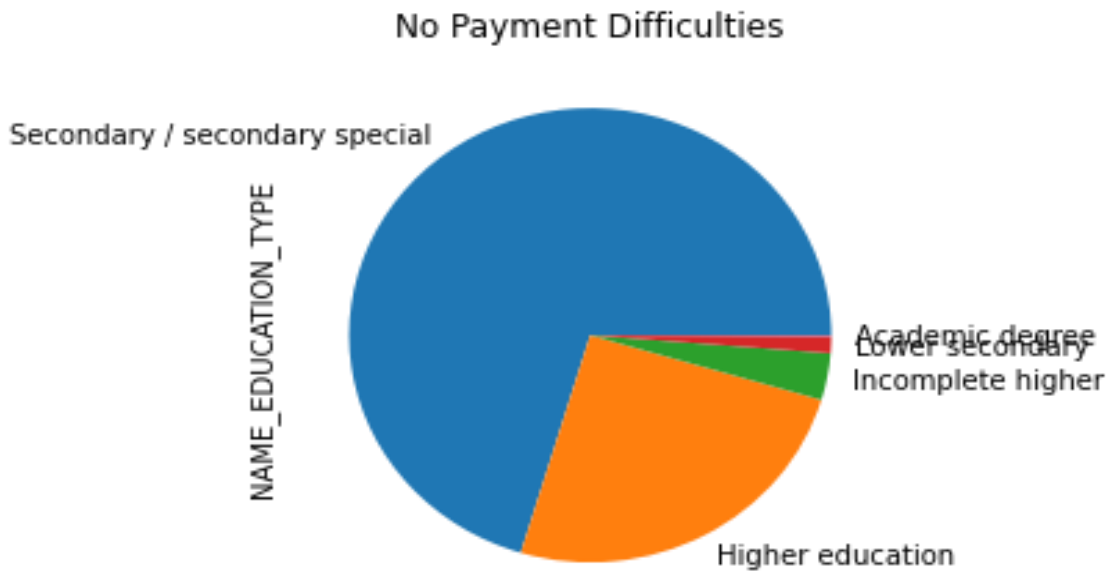
After observing both the graphs we conclude the following:

1. Working people is paying on time which is generating maximum revenue for bank and at the same time they are the one's not paying on time, which is approximately same and shows no profit no loss term.
2. Commercial associates are likely to cause less payment difficulties which results in profit to the bank.
3. As per the comparison of pensioners more no. of pensioners pay on time compared to those who not. Same goes for State servants as well.
4. We can also say Businessman are taking less loan but always pay on time.
5. Even Students pay on time considering their parents pay the money also students tend to take less loans, as they are not earning.
6. Unemployed and Maternity leave people are same for causing payment difficulties and no payment difficulties.



EDUCATION TYPE

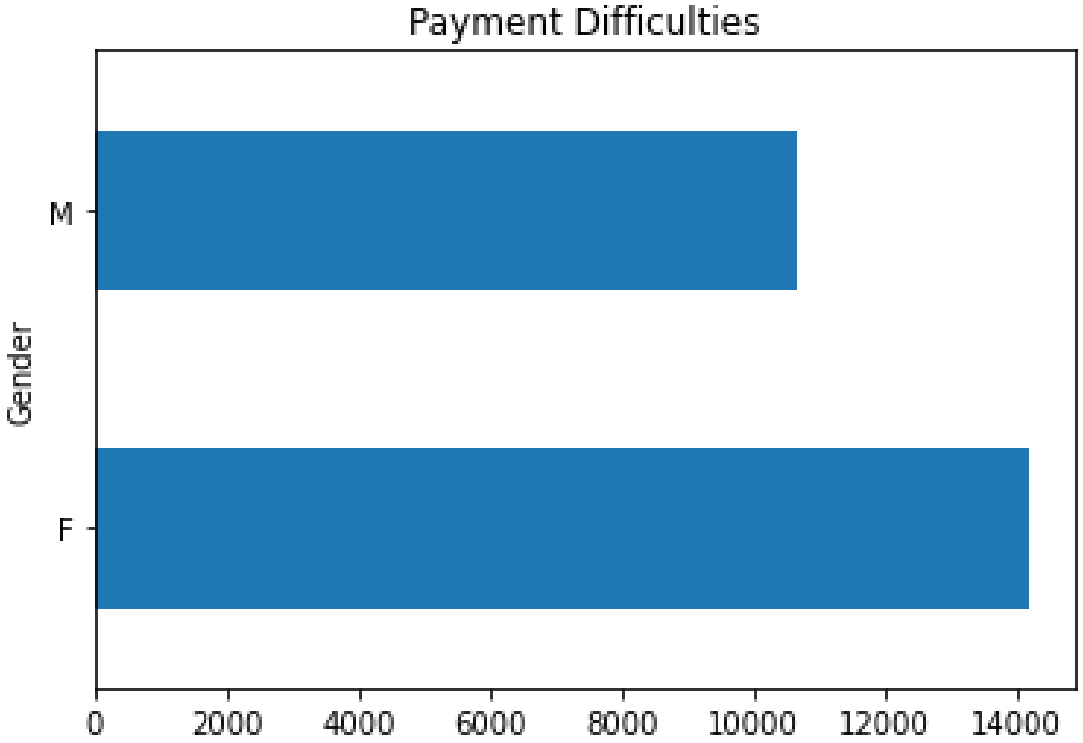


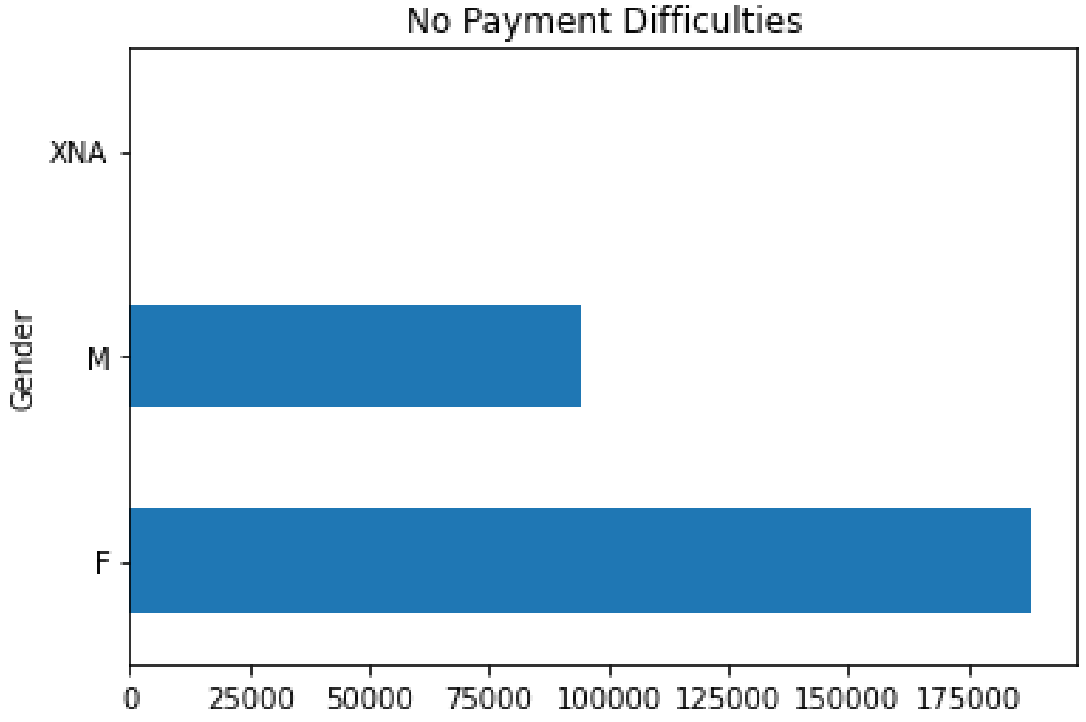


From the above pie chart we conclude the following:

- Secondary Education people tend to take more loans.
- At the same time Secondary Education type make more payment difficulties
- Academic degree people tend to take less loans alternatively they tend to cause less payment difficulties to bank.
- Higher Education people are tend to cause less payment difficulties considering they are highly educated.

GENDER

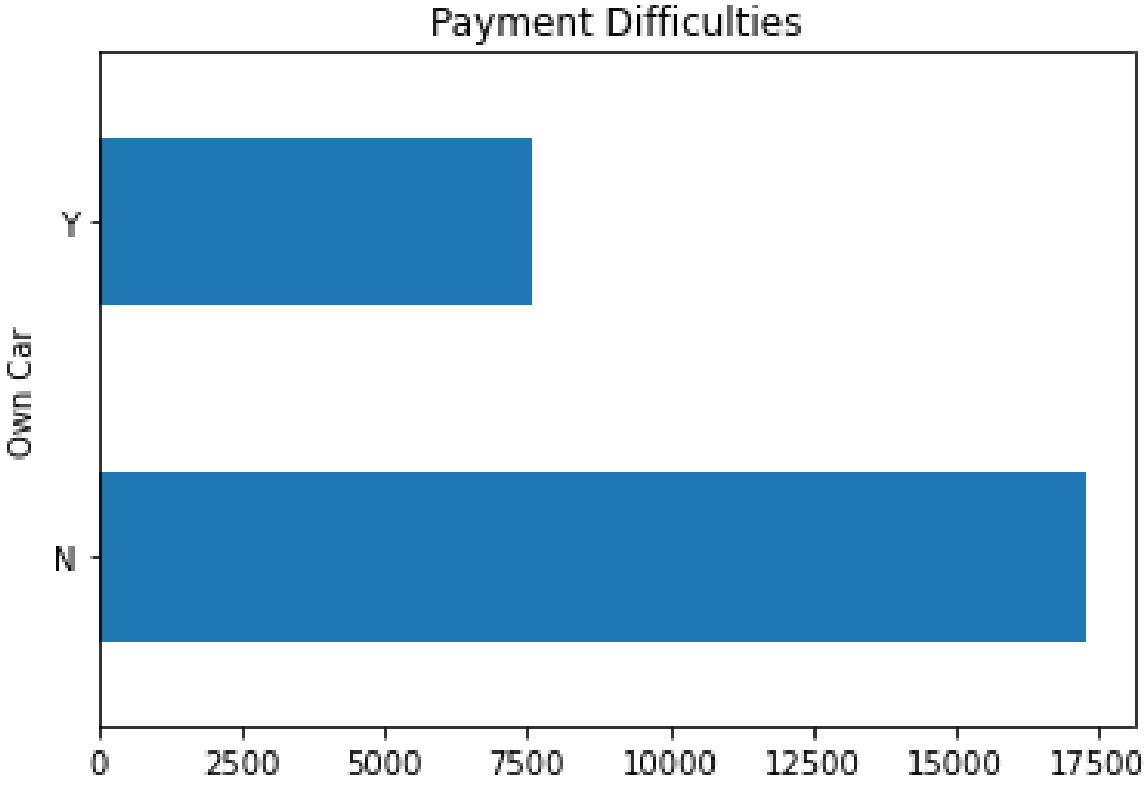


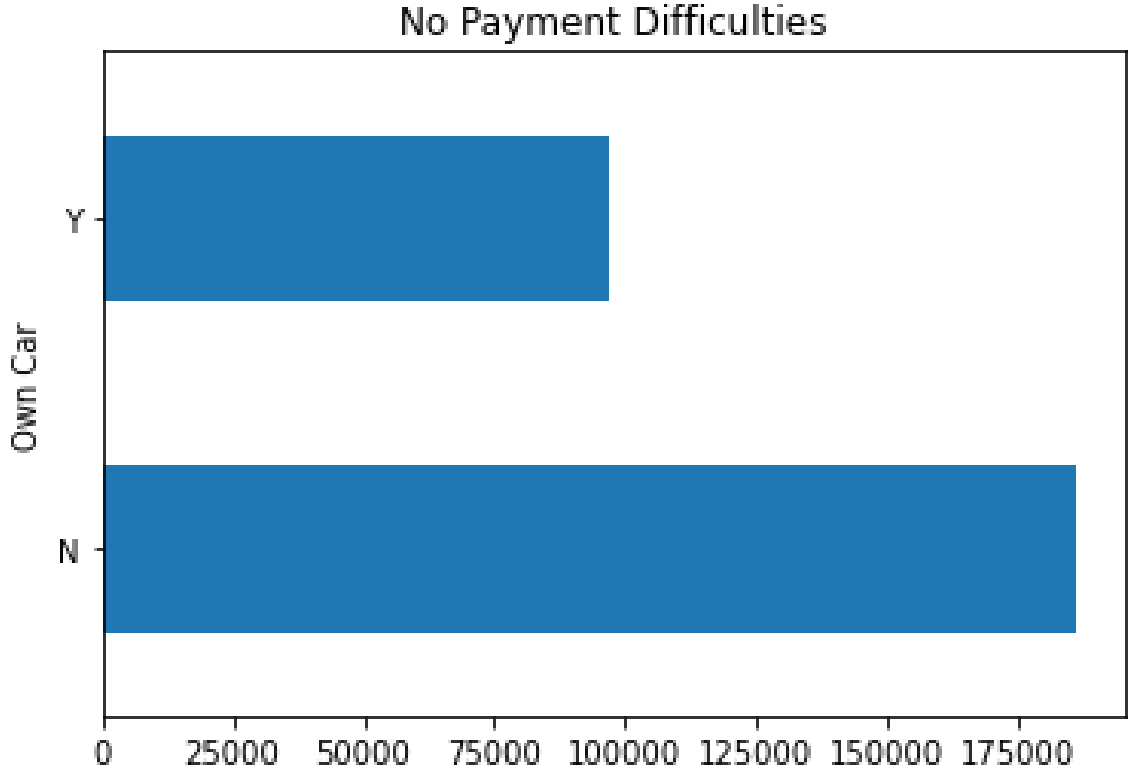


After observing the above two plots we conclude the following:

- Female tend to take more loans then men.
- But payment difficulties caused by Female is approx 14,000 compared to those who do not cause is approx 2,00,000.
- For Males, out of approx 1,20,000/- only 11,000/- males cause payment difficulties.
- We can say that in both the cases only 8-9% are likely to default in payment.
- XNA is info that is not available where we can assume that people don’t prefer to say their gender.

OWNS A CAR

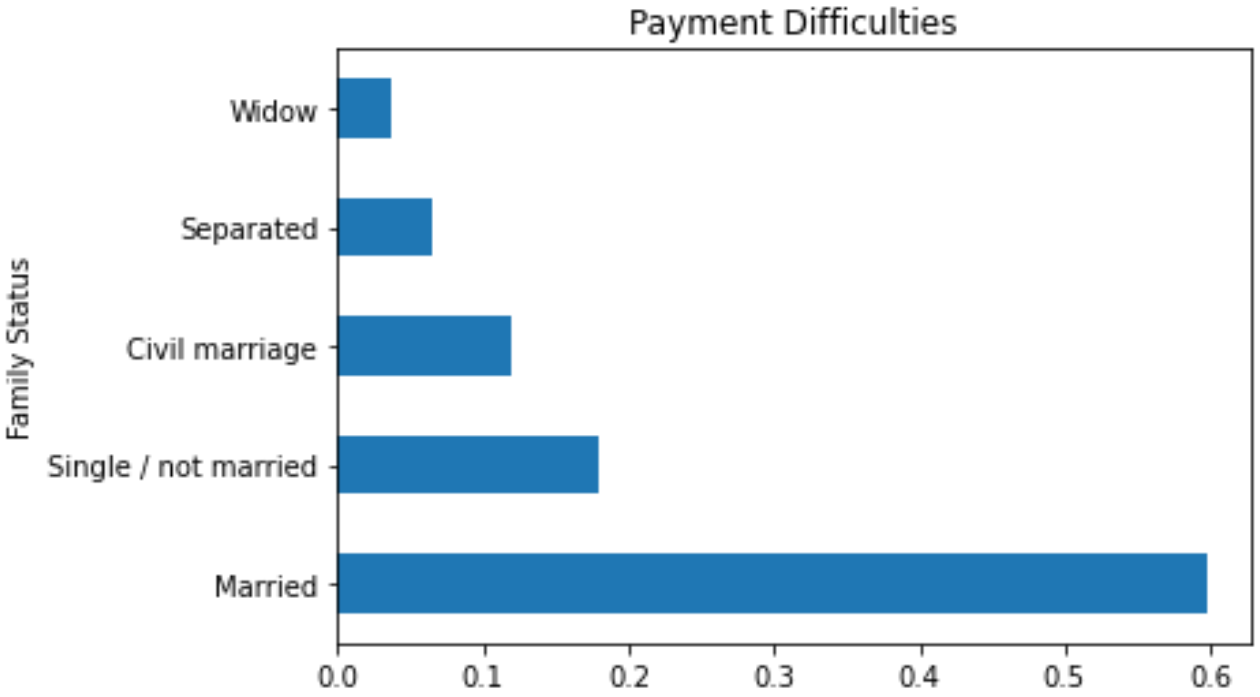


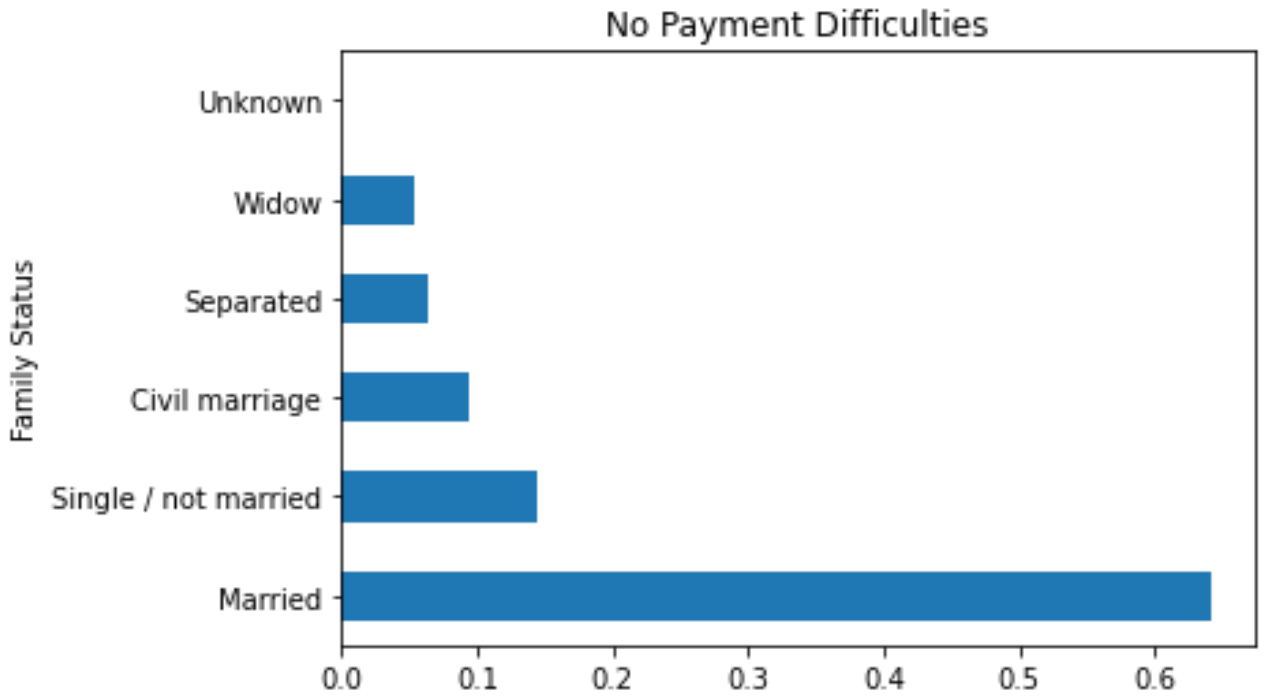


After observing the above two plots we conclude the following:

- People who don't own a car take more loans compared to those who own a car.
- Also people who don't own a car cause more payment difficulties to bank than those who own a car.

FAMILY STATUS

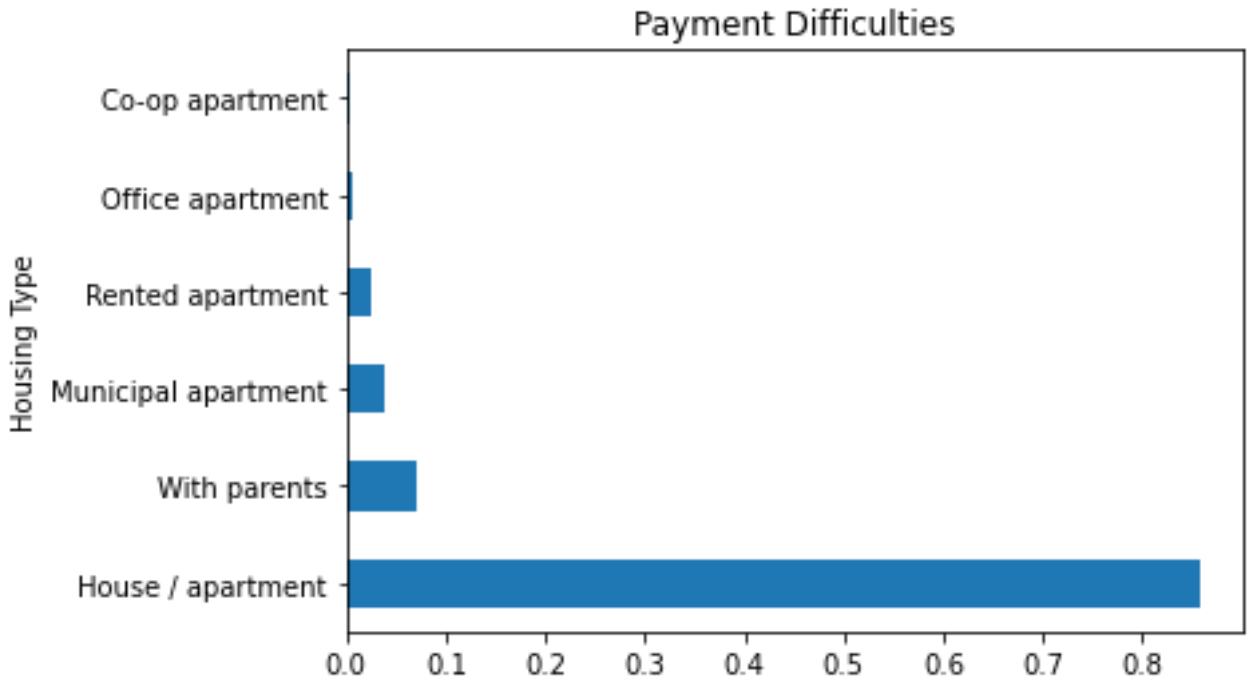


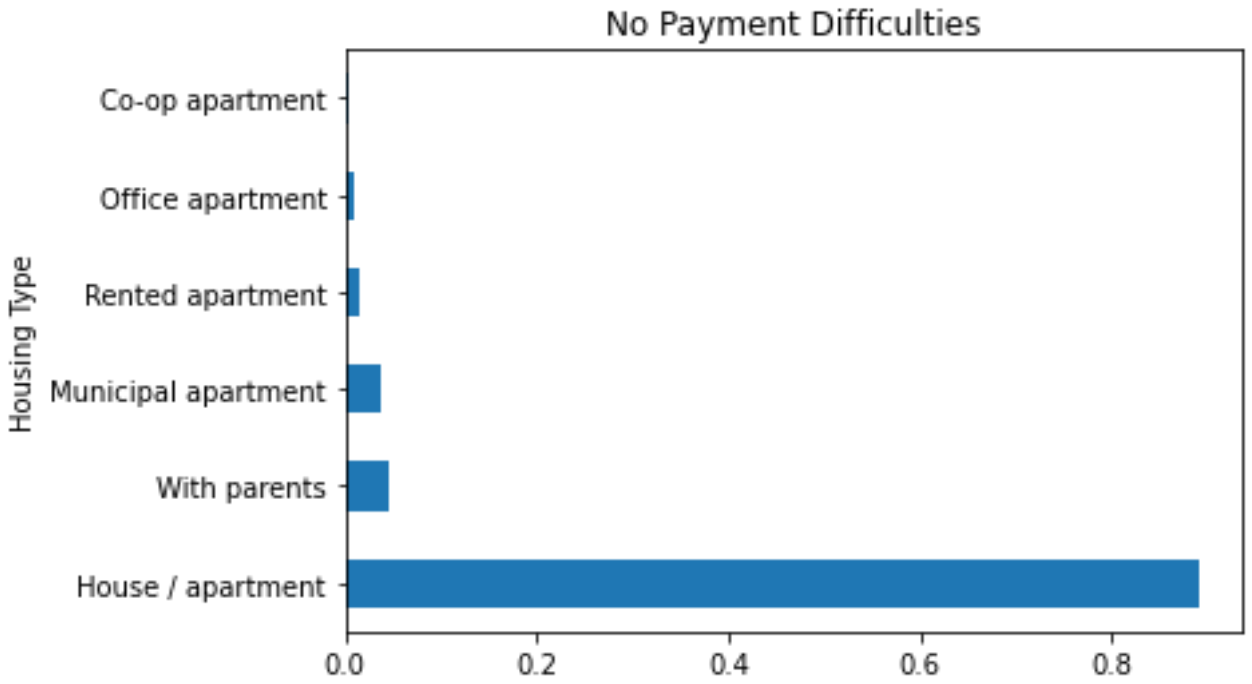


After observing the above two plots we conclude the following:

- Married people take more loans followed by Single/not married, Civil Marriage, Separated and Widow.
- There are some unknown values in No Payment Difficulties which can be ignored, as it doesn't effect our analysis.
- Approx all the values in the graph is same which concludes that the family status does impact the bank.

HOUSING TYPE

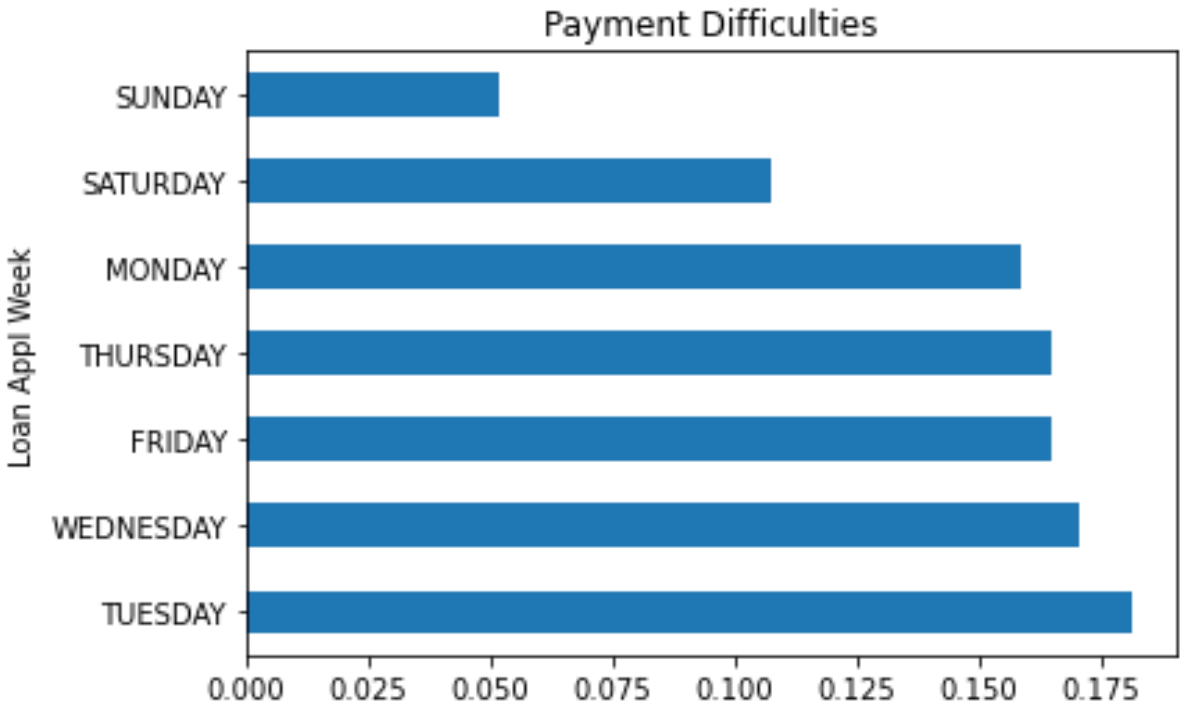


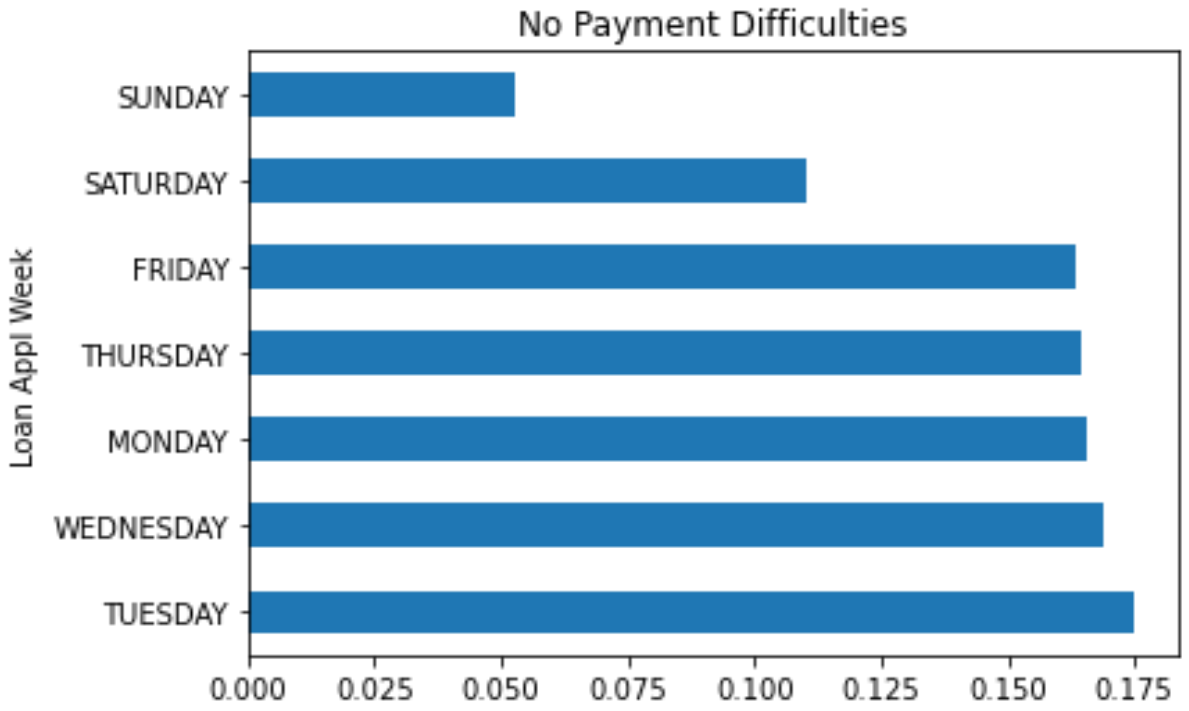


After observing the above two plots we conclude the following:

- Maximum number of loans applicant are having their own house.
- Also we can see some even stay in rented apartment, where they are causing payment difficulties, considering the reason every month they even have to pay their House Rent. Giving Loans to Rented Apartment is little bit risky.
- As per the people who own a house is neutral in both the graphs.

LOAN APPLICATION IN A WEEK

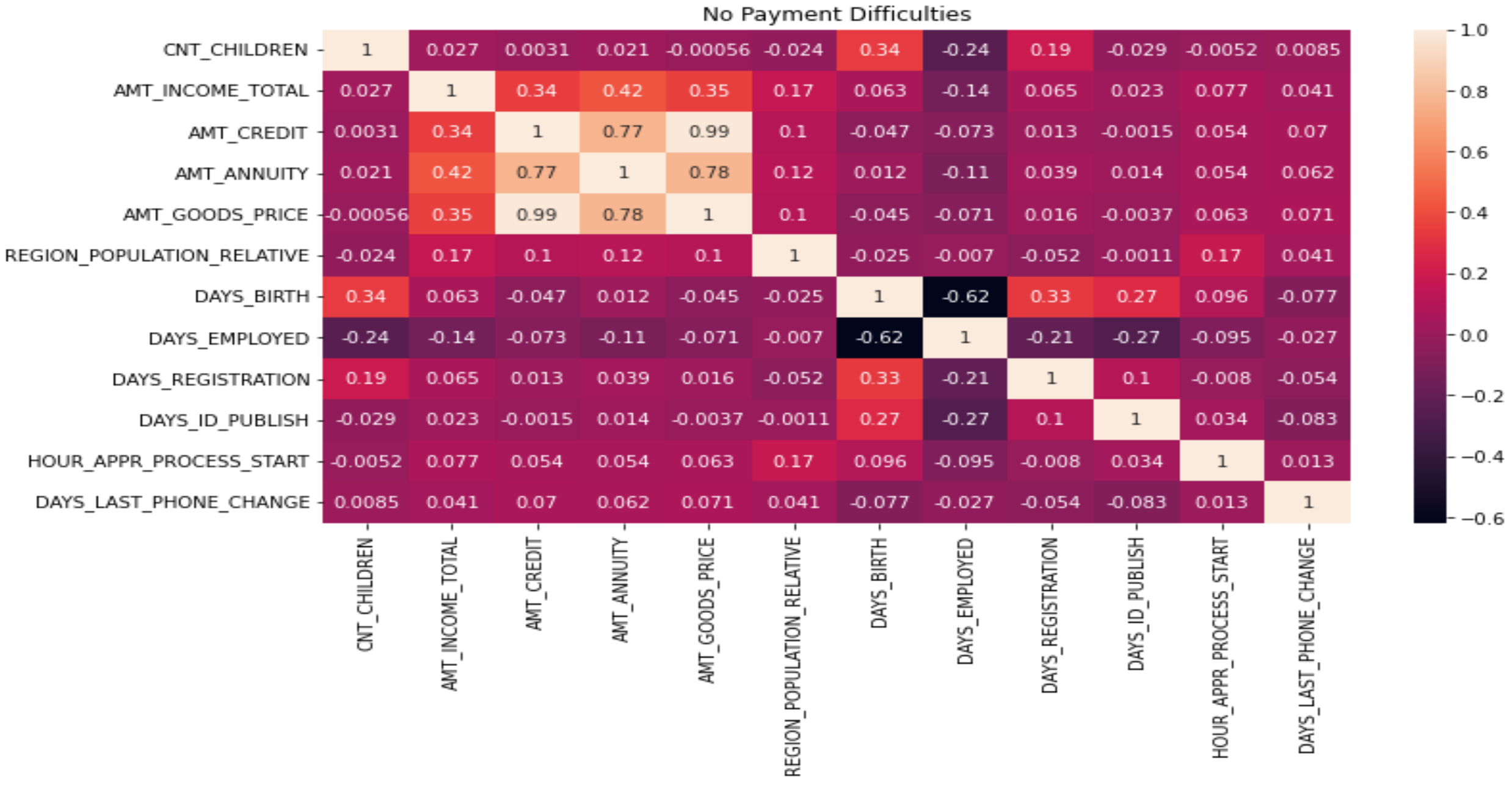




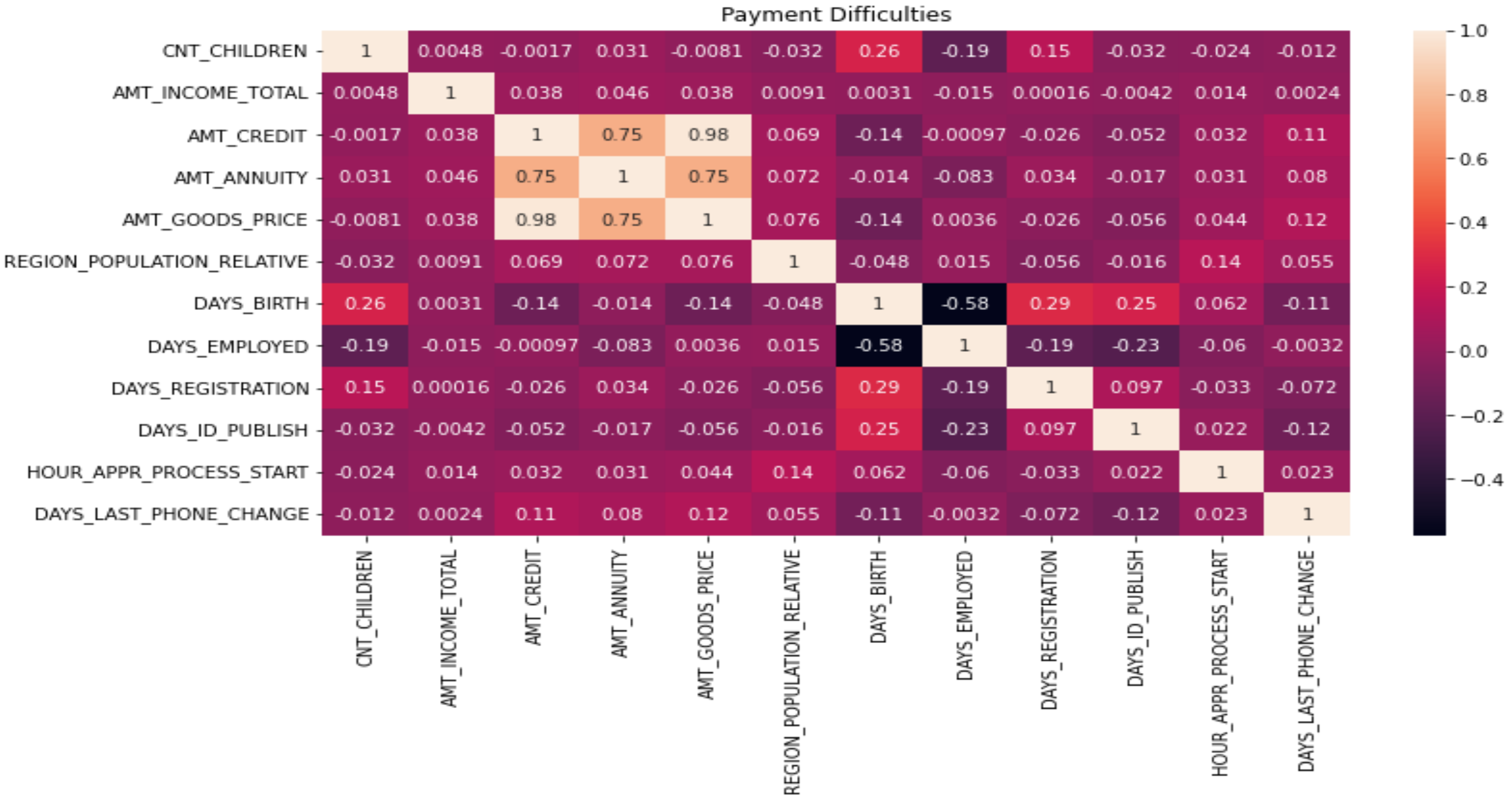
Above two graphs conclude the following:

- Maximum no. of loan applicants starts on Tuesday and Sunday.
- Whereas on Saturday and Sunday least no. of applications are there.

COORELATION FOR TARGET 0



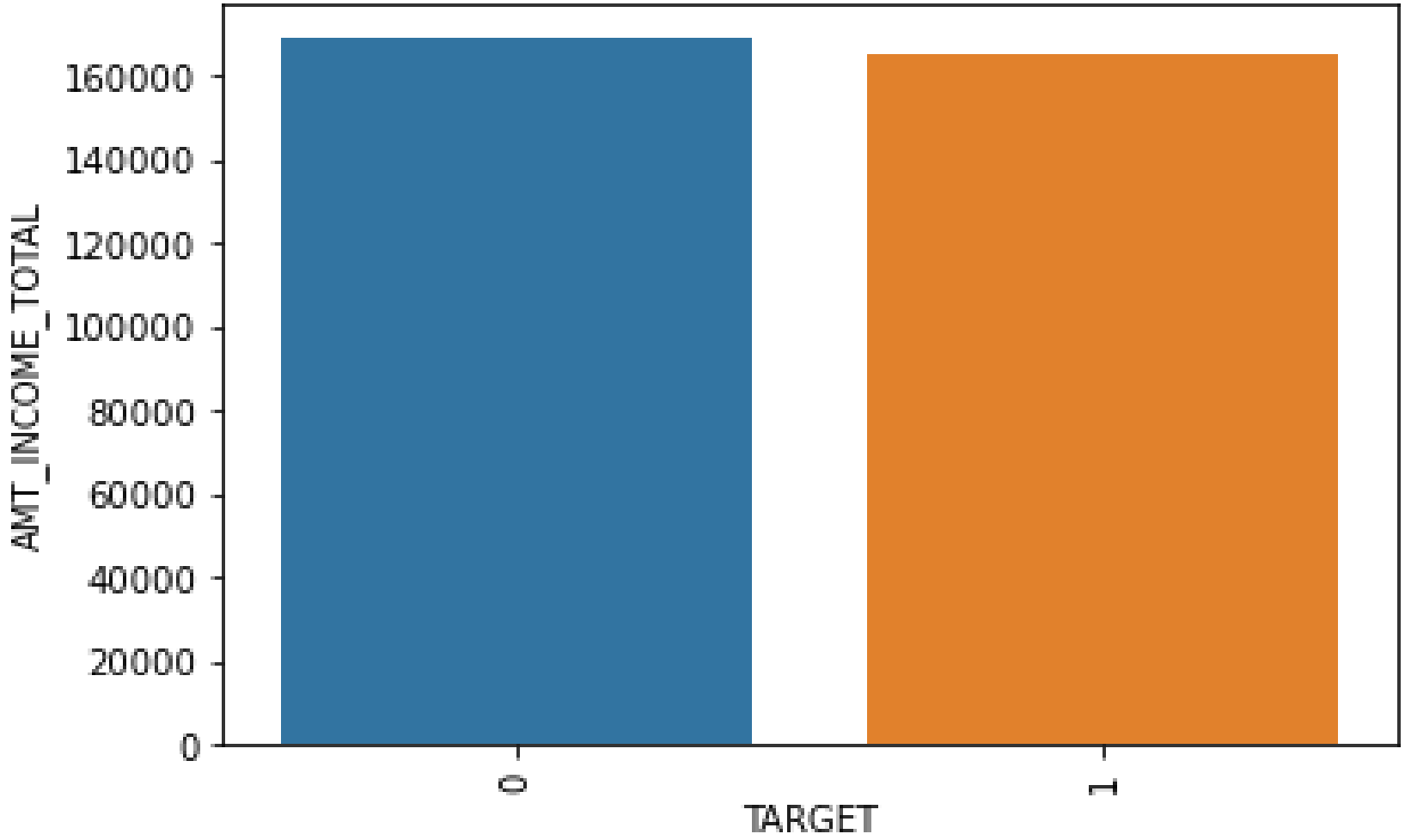
COORELATION FOR TARGET 1



TARGET AND INCOME

From the graph we conclude that:

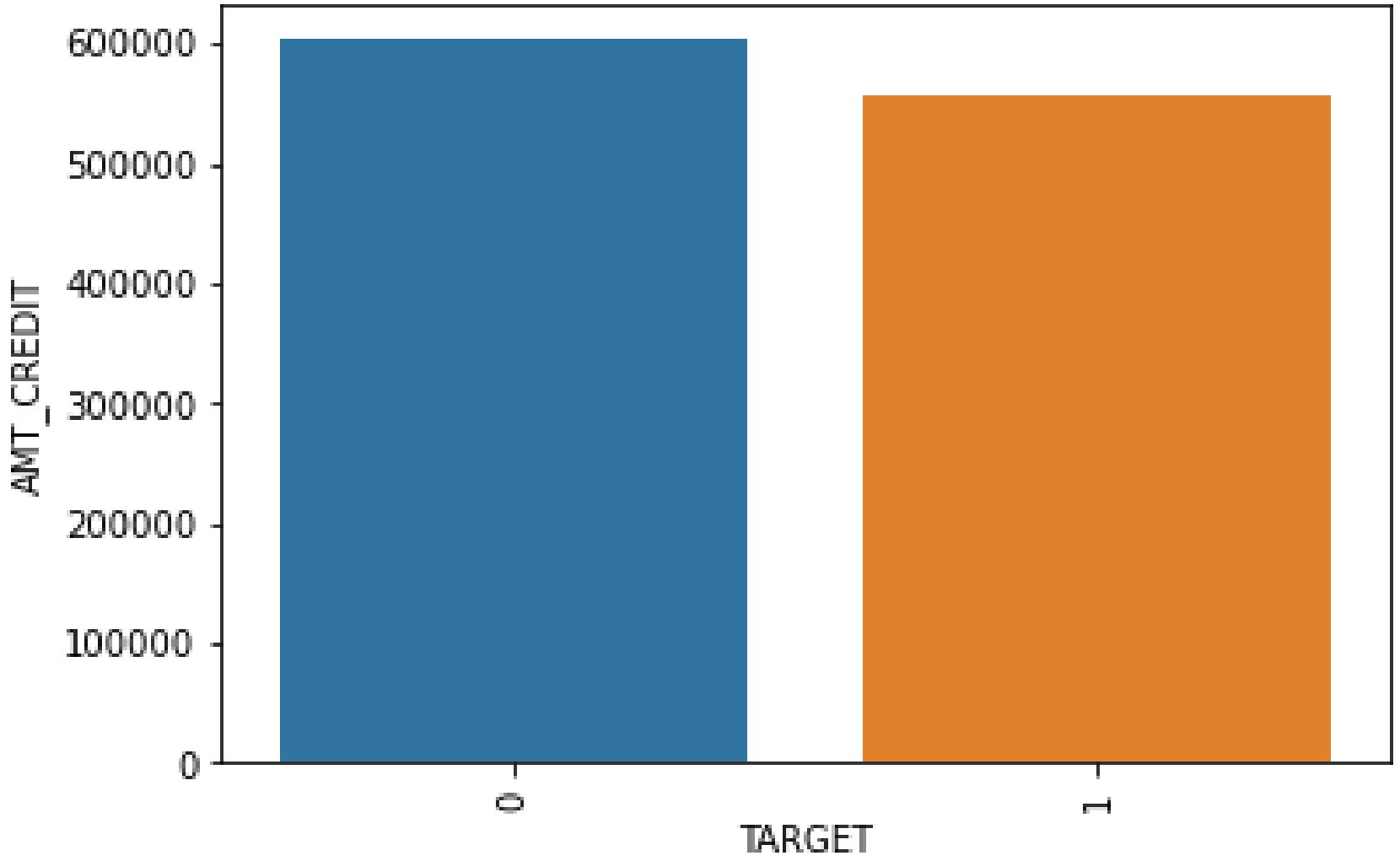
- Target 0 is people with No Payment Difficulties and Target 1 is people with Payment Difficulties.
- Target 0 people's total income is more than Target 1.
- Also Target 0 people don't default on any payments as their income level is high.



TARGET AND CREDIT AMOUNT

From the graph we conclude that:

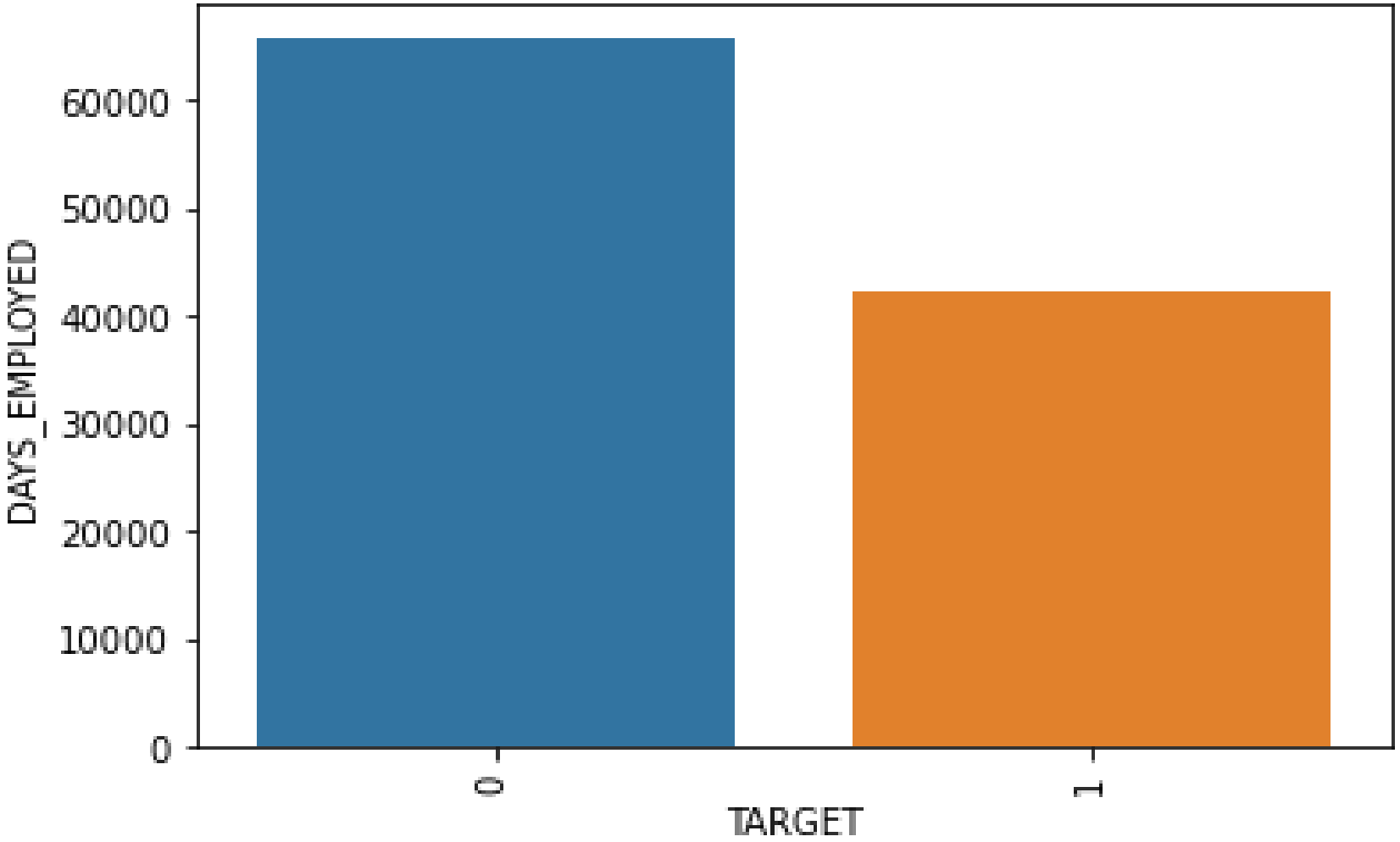
- Target 0 is people with No Payment Difficulties and Target 1 is people with Payment Difficulties.
- Target 0 people's credit amount is more than Target 1.
- With this we can conclude people with higher credit does not default on payment.



TARGET AND DAYS EMPLOYED

From the graph we conclude that:

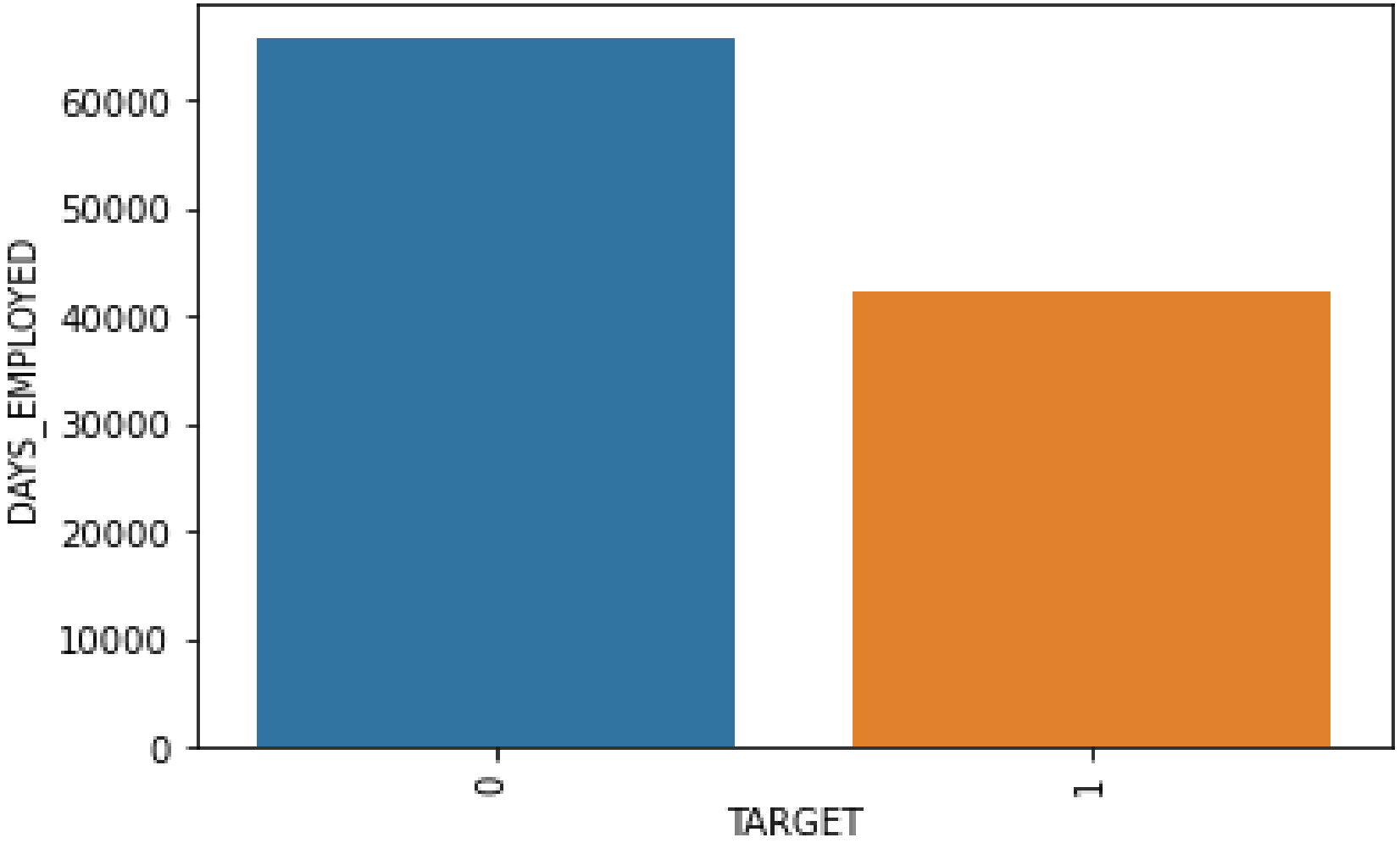
- Target 0 is people with No Payment Difficulties and Target 1 is people with Payment Difficulties.
- Target 0 people at the time of applying for the loan were already working for a long period of time compared to Target 1.



TARGET AND DAYS EMPLOYED

From the graph we conclude that:

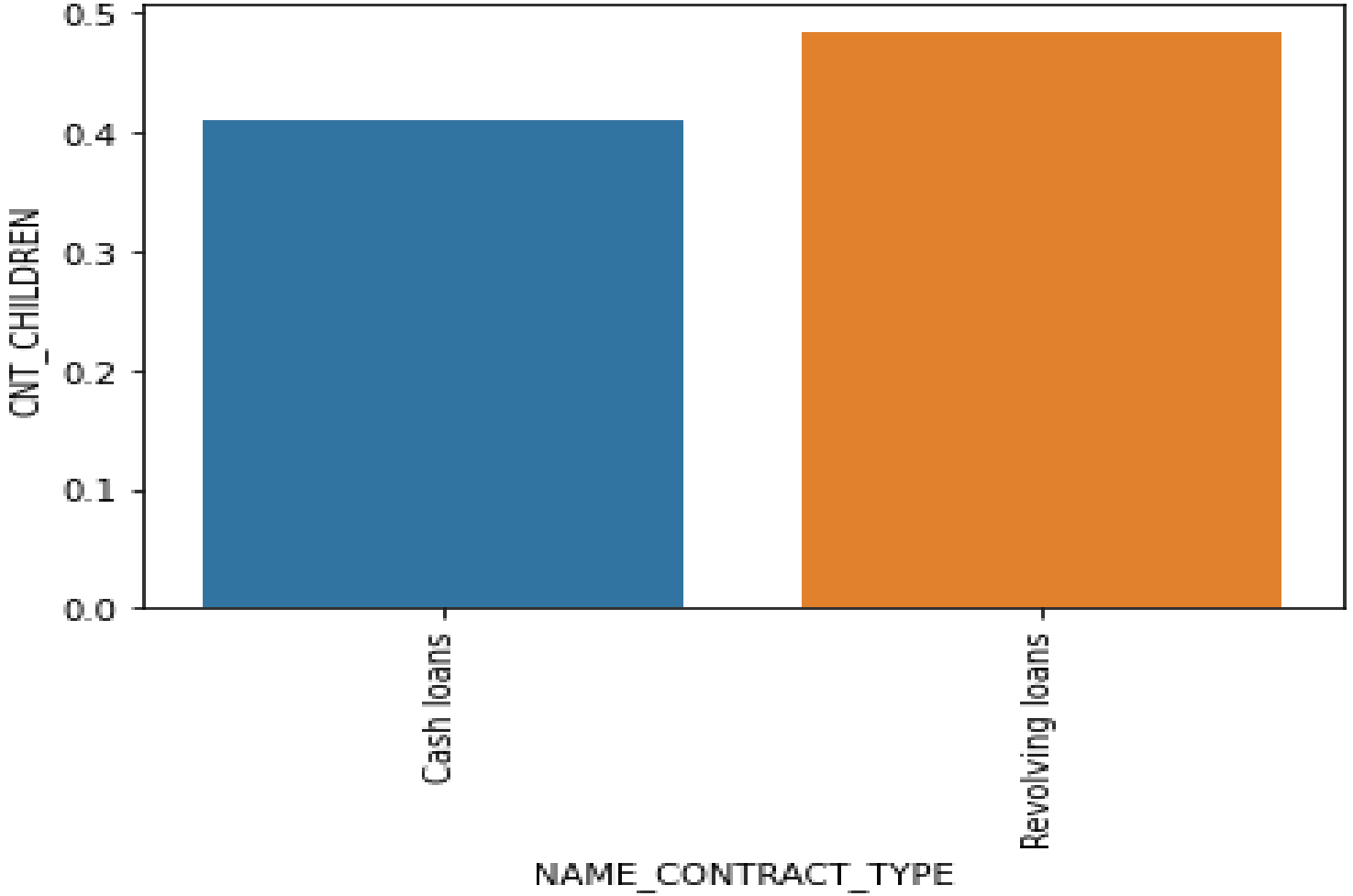
- Target 0 is people with No Payment Difficulties and Target 1 is people with Payment Difficulties.
- Target 0 people at the time of applying for the loan were already working for a long period of time compared to Target 1.



CONTRACT TYPE AND NO. OF CHILDREN

From the graph we conclude that:

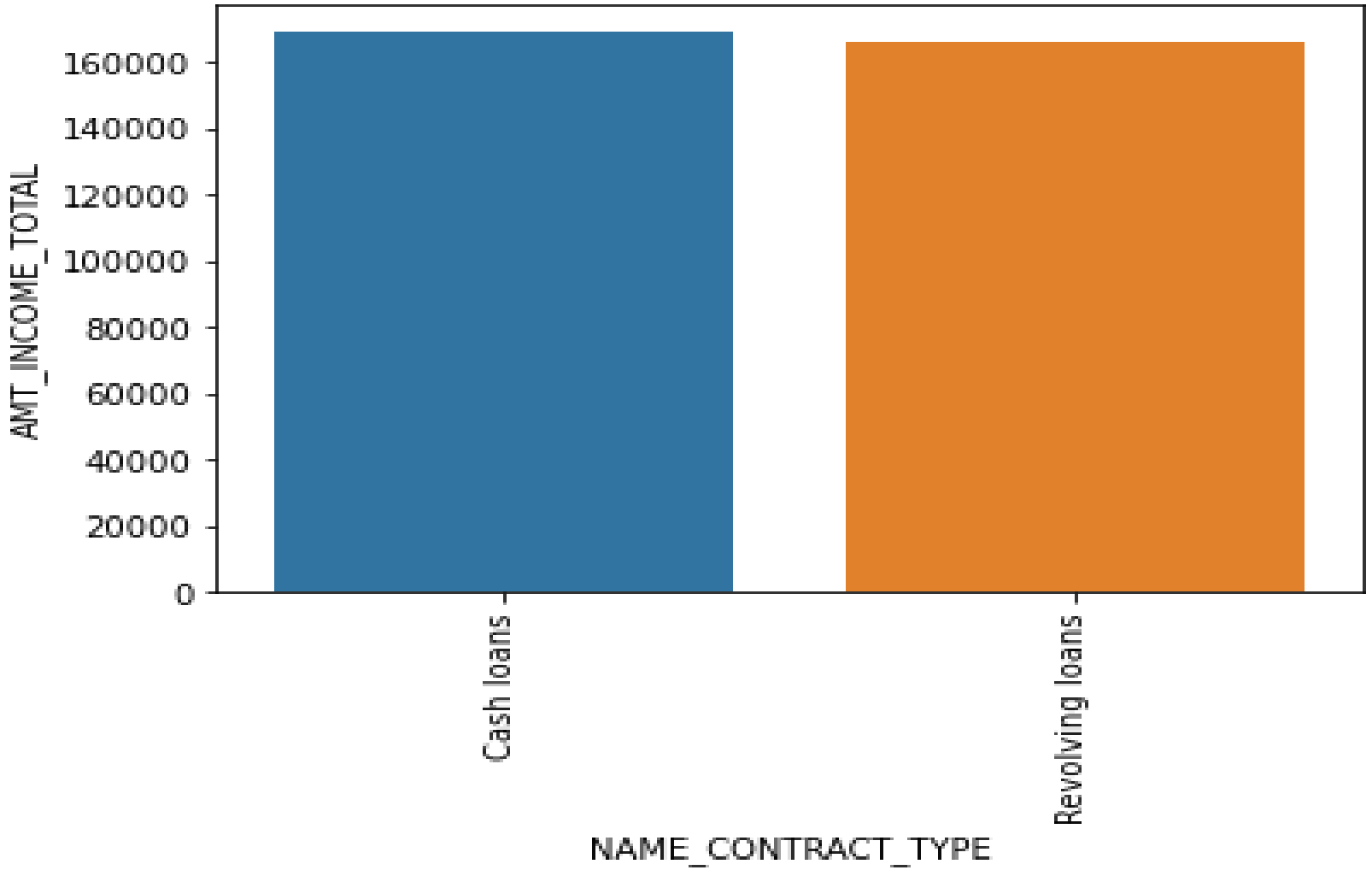
- Revolving loan clients have more children than cash loans.
- Considering the more no. of children the more loans they require.



TOTAL INCOME AND CONTRACT TYPE

From the graph we conclude that:

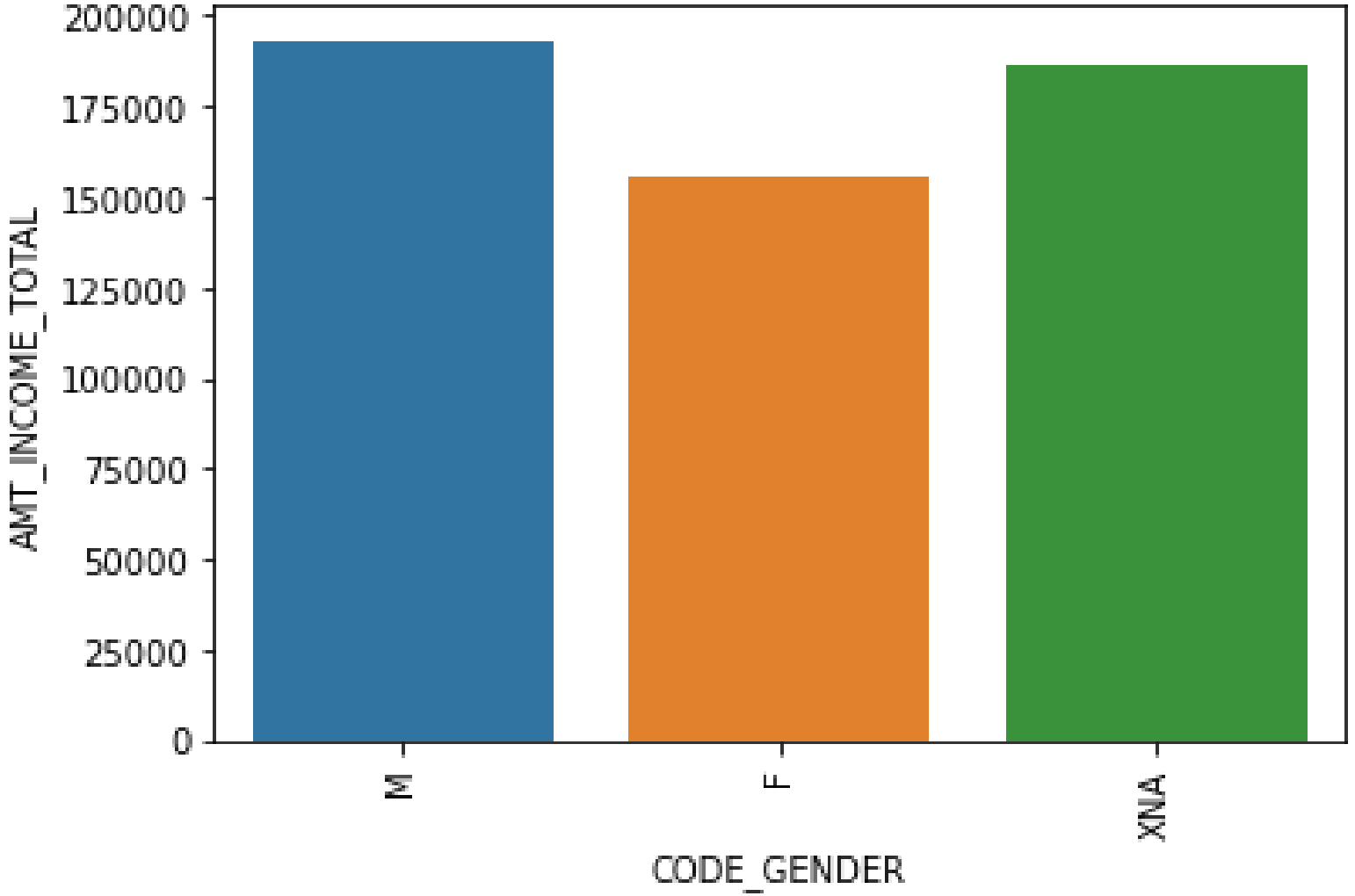
- Higher income people tend to take more cash loans compared to revolving loans.



TOTAL INCOME AND GENDER

From the graph we conclude that:

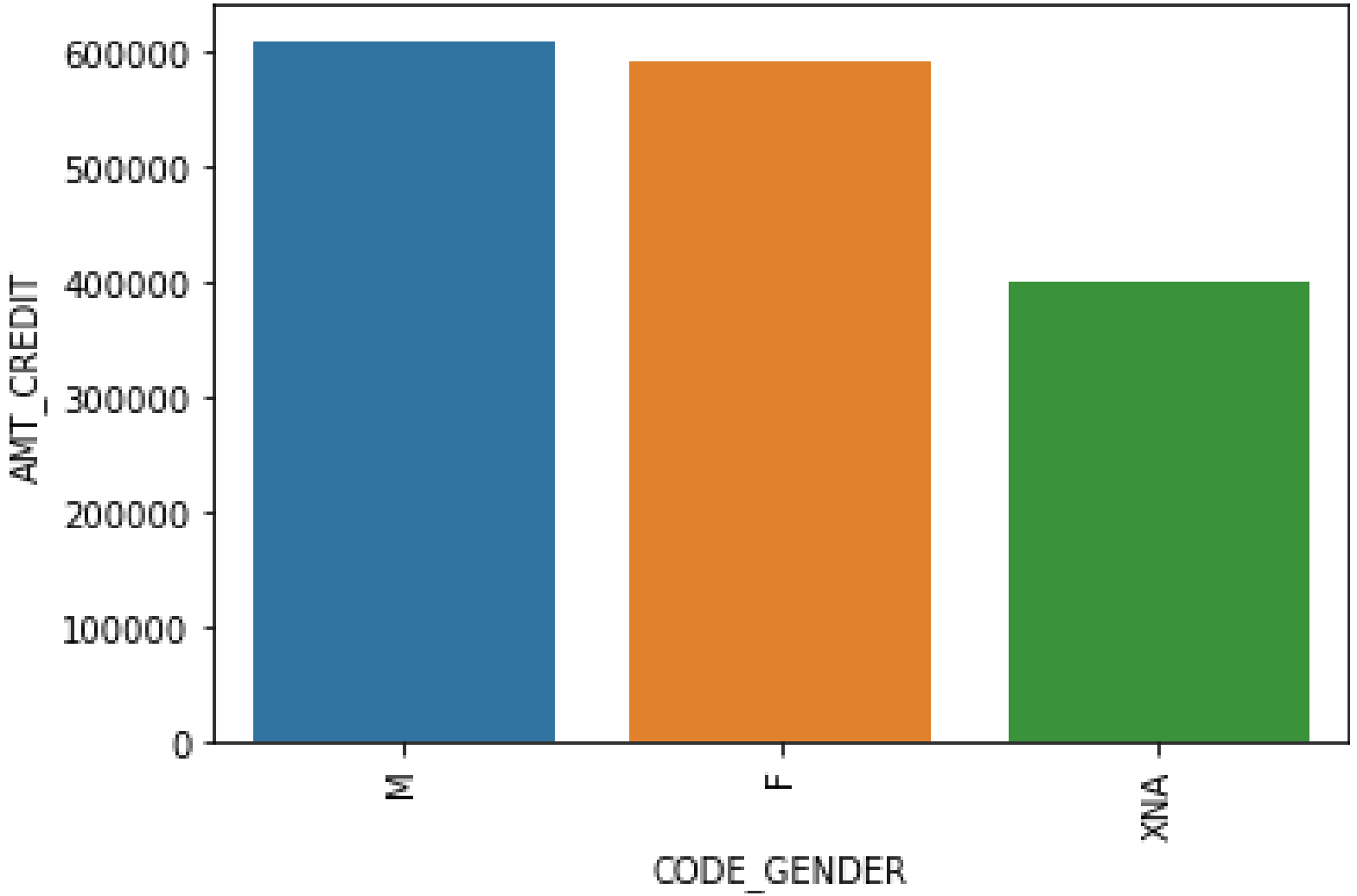
- Male clients have more income compared to Female clients.
- XNA is gender not available considering client doesn't prefer to mention their gender.
- XNA Clients total income is more compared to Female Clients



CREDIT AMOUNT AND GENDER

From the graph we conclude that:

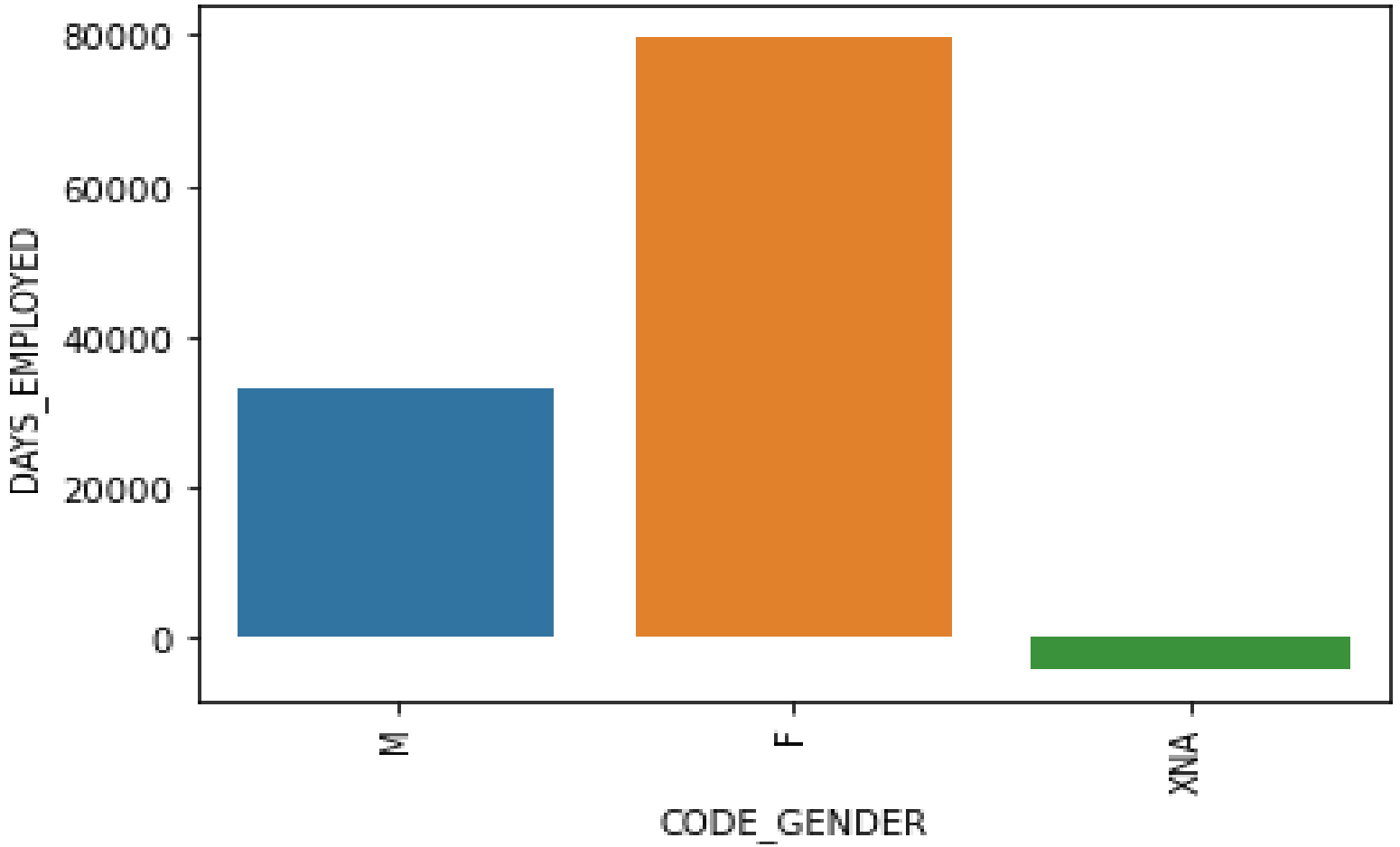
- Credit amount of Male Clients is more than Female Clients.



DAYS EMPLOYED AND GENDER

From the graph we conclude that:

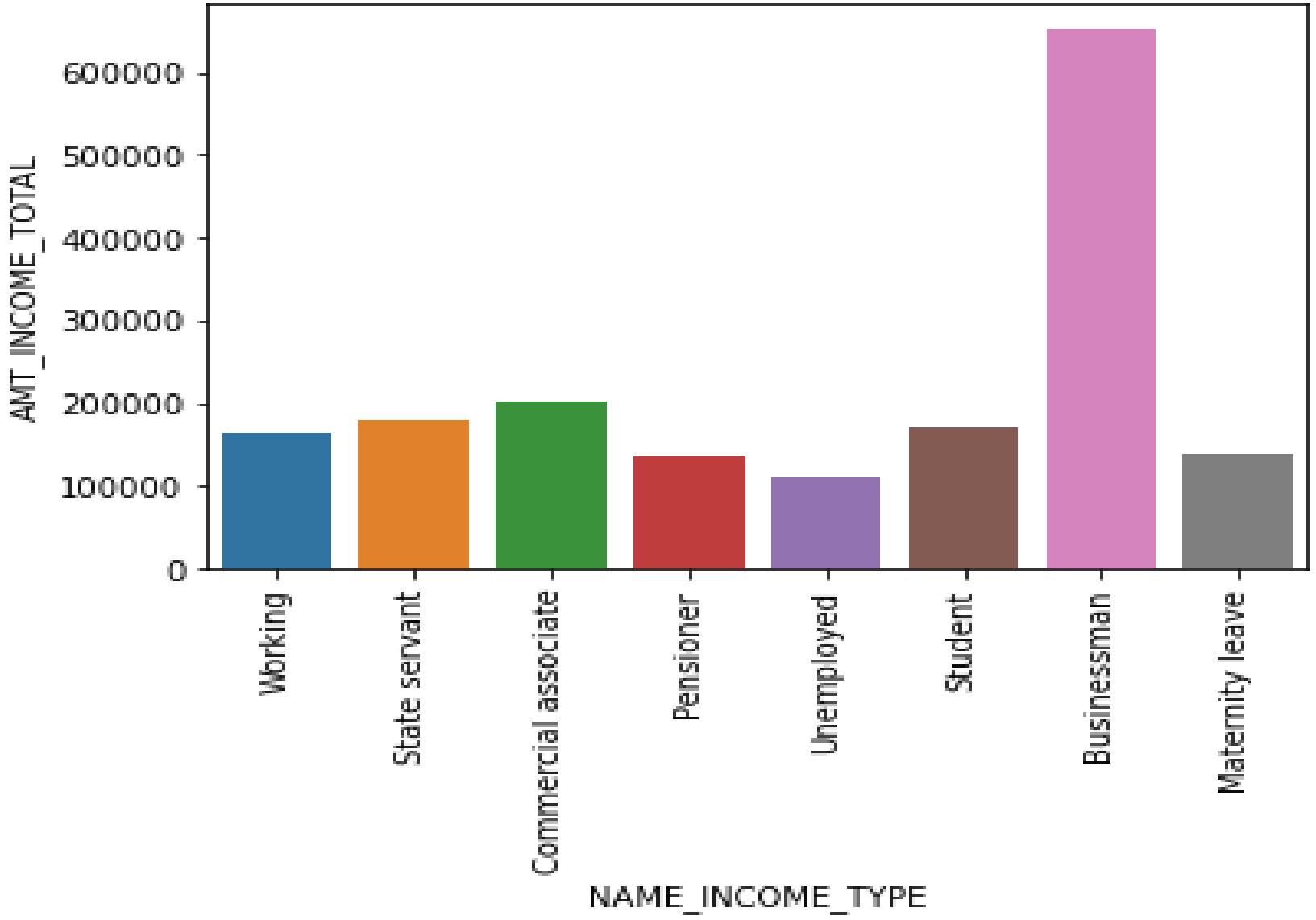
- Female clients are employed for more days at the time of application, compared to male.



TOTAL INCOME AND INCOME TYPE

From the graph we conclude that:

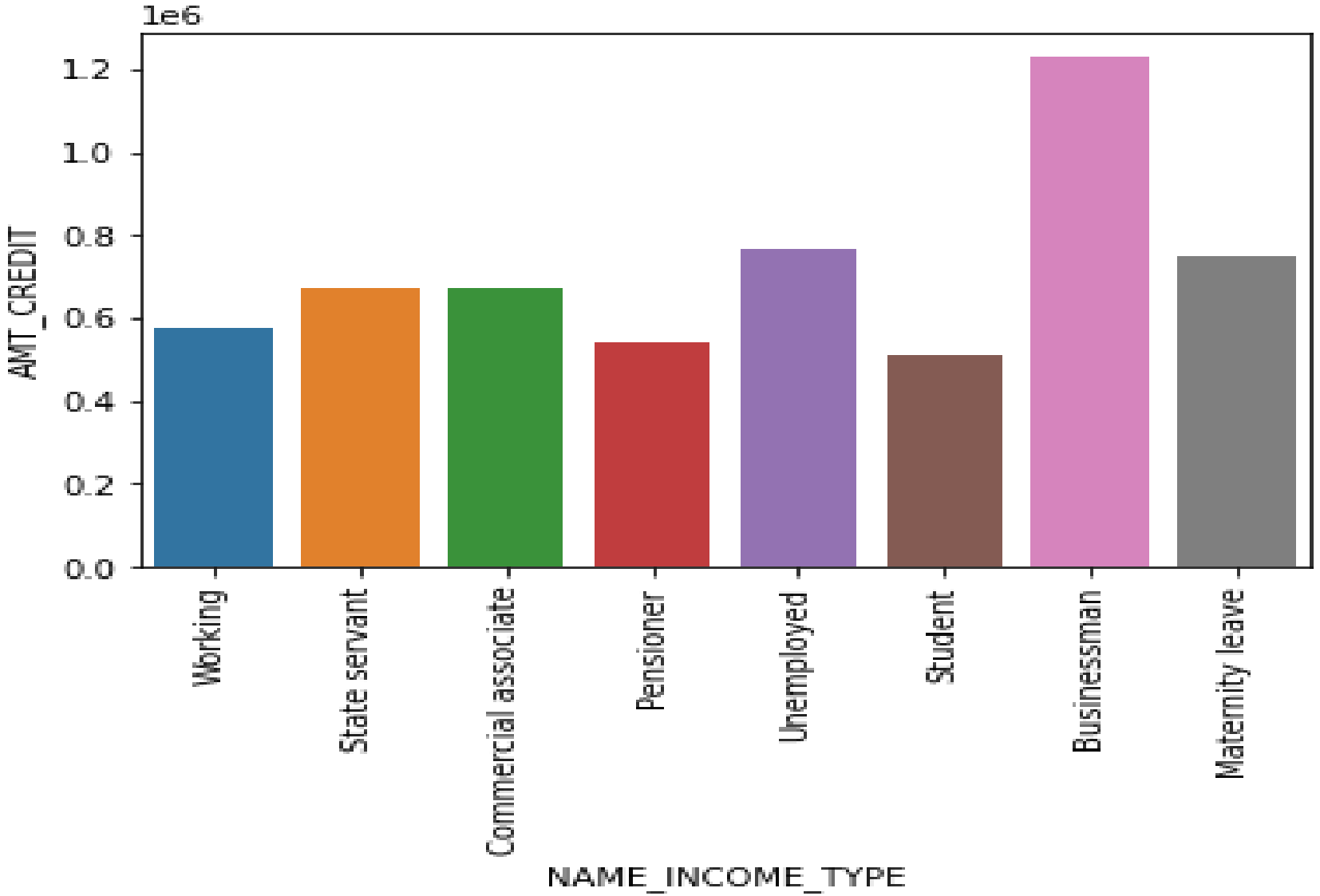
- Income type of businessman is the highest.
- Income type of unemployed is the lowest.



CREDIT AMOUNT AND INCOME TYPE

From the graph we conclude that:

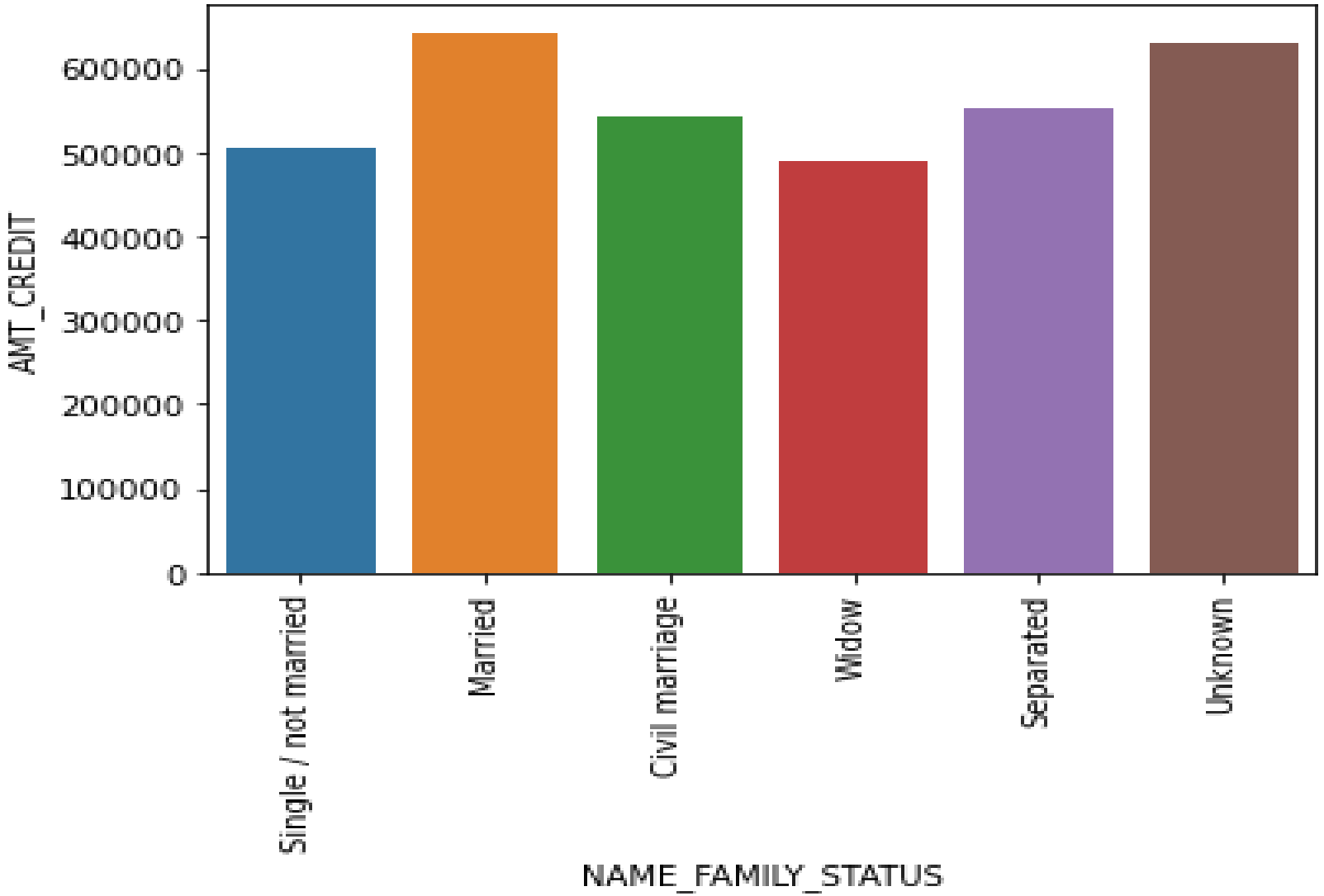
- Credit amount of businessman is the highest.
- Credit amount of student is the lowest.



CREDIT AMOUNT AND FAMILY STATUS

From the graph we conclude that:

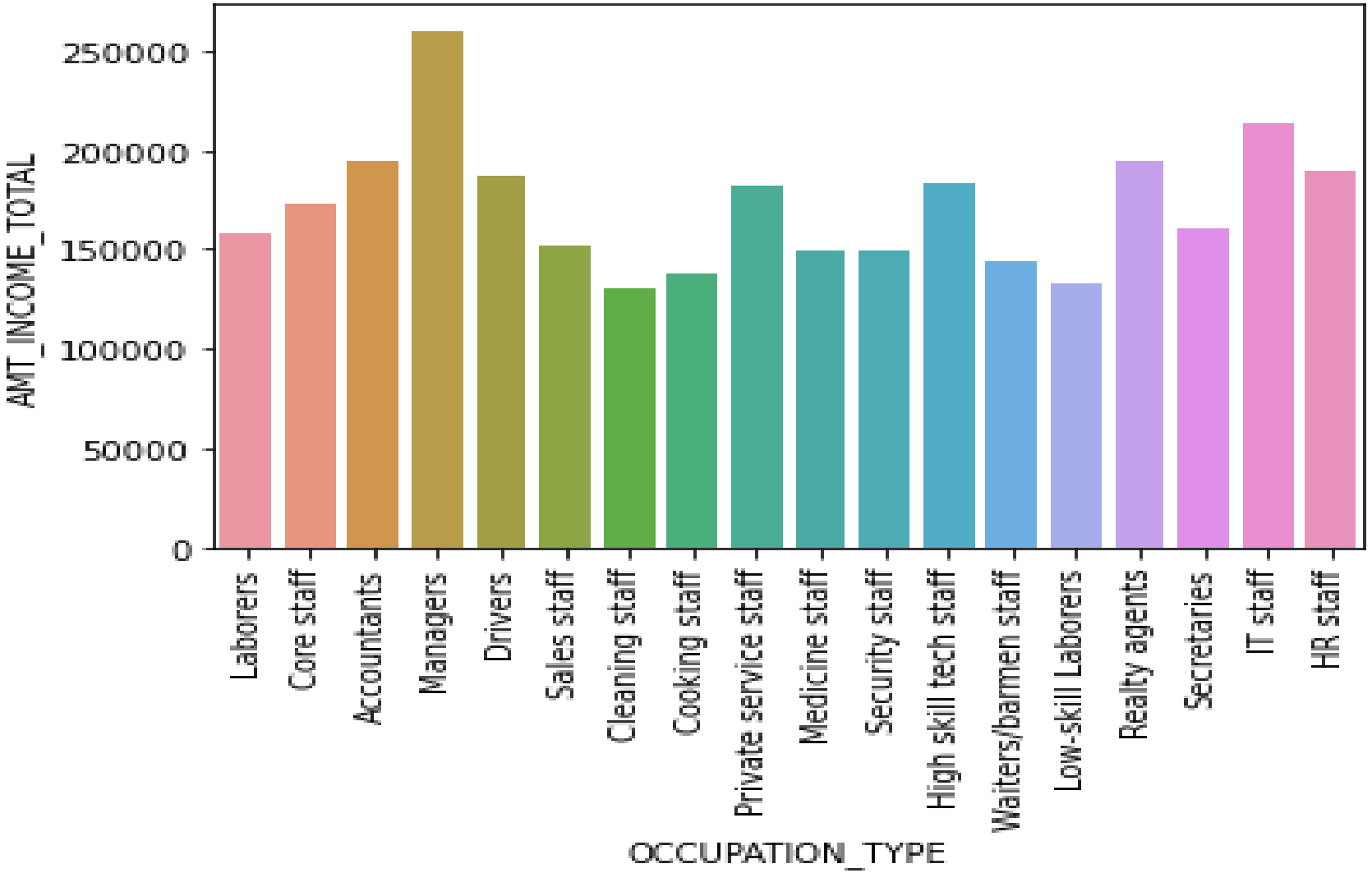
- Married Clients have the highest credit amount.
- Widow Clients have the lowest credit amount.



TOTAL INCOME AND OCCUPATION TYPE

From the graph we conclude that:

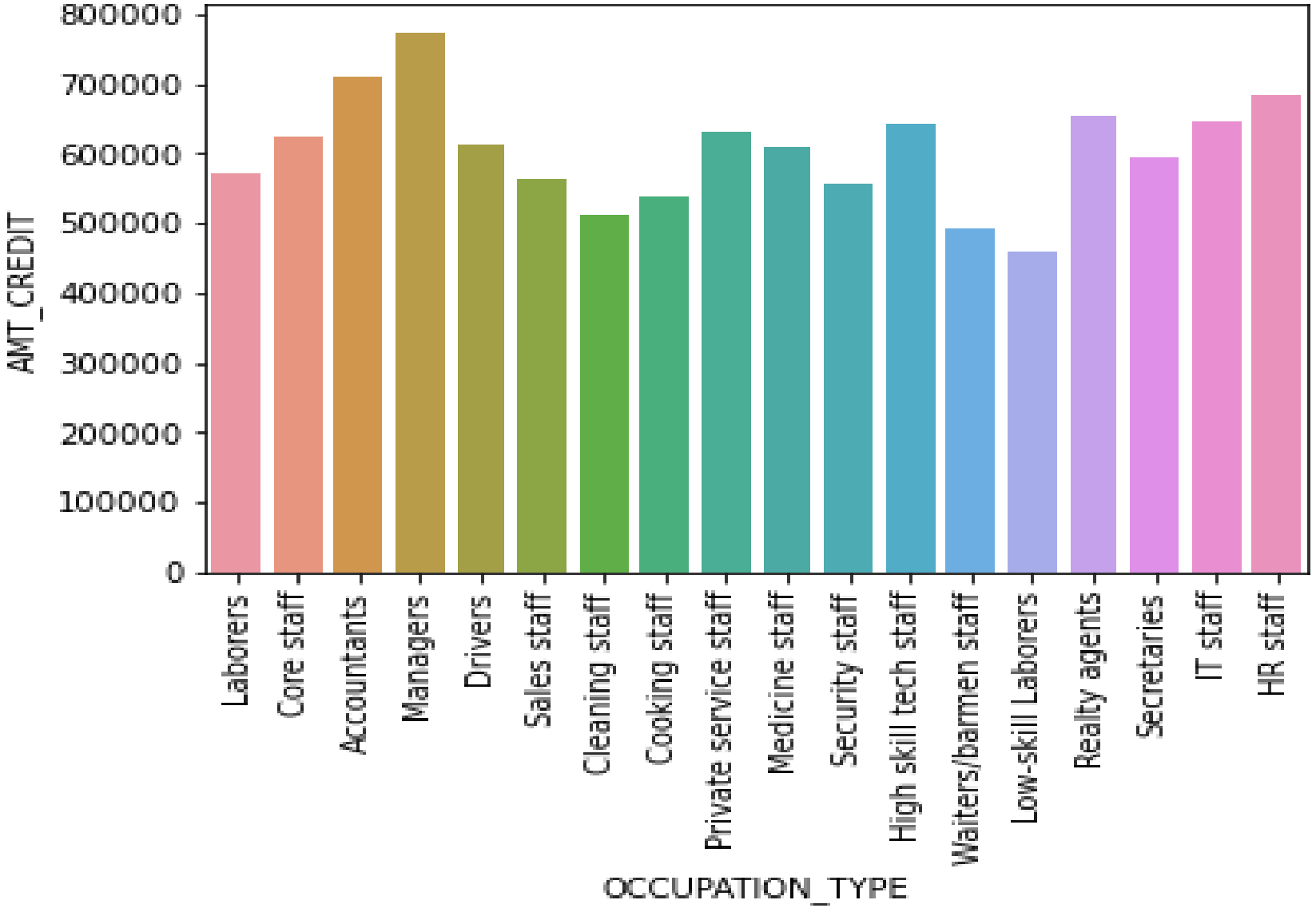
- Income of Managers are maximum compared to the rest of the Occupation.
- Cleaning Skill and Low-Skill Laborers have the lowest income.



CREDIT AMOUNT AND OCCUPATION TYPE

From the graph we conclude that:

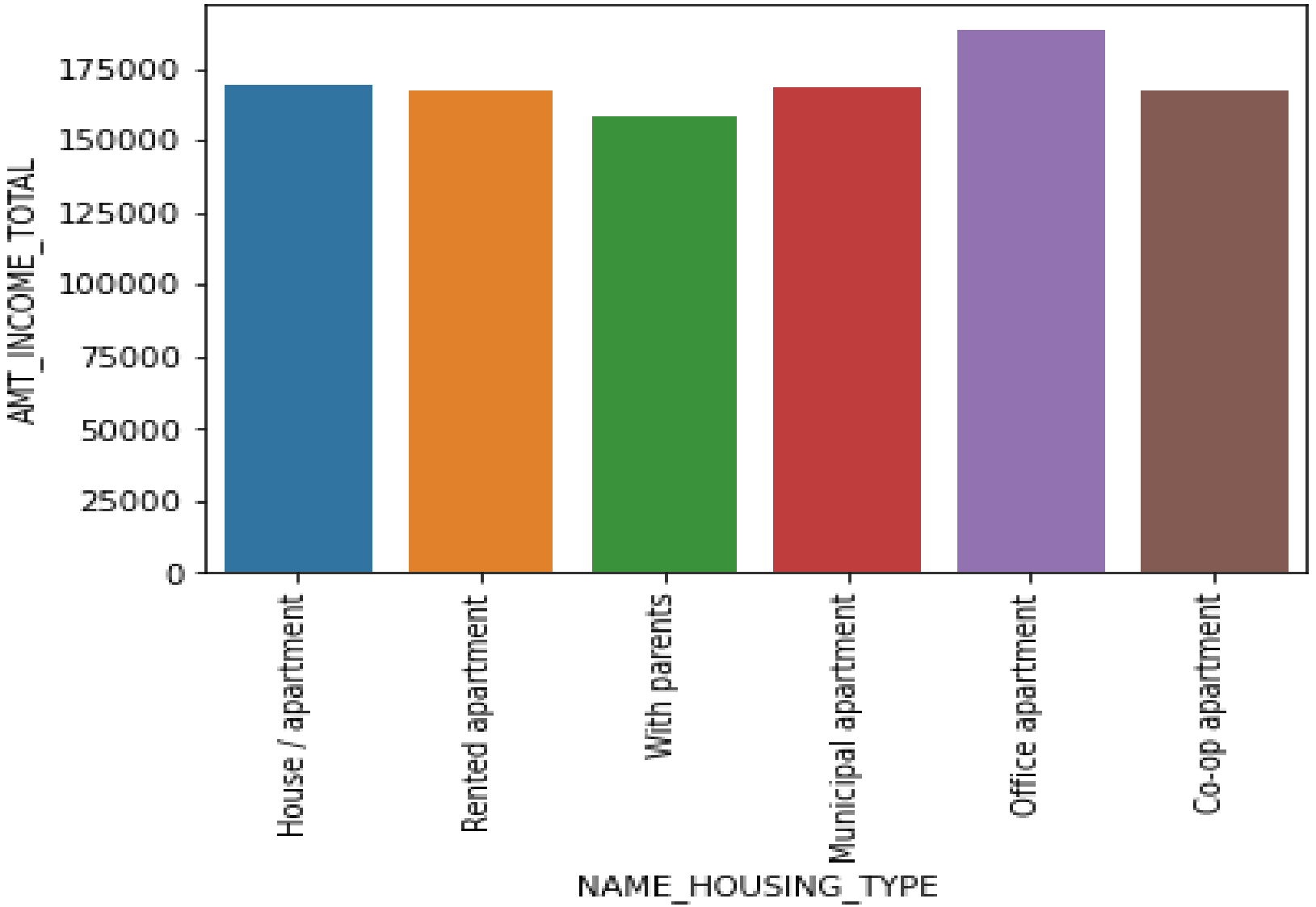
- Managers have the highest credit amount, followed by Accountants.
- Low-Skill Laborers have the lowest credit amount.



TOTAL INCOME AND HOUSING TYPE

From the graph we conclude that:

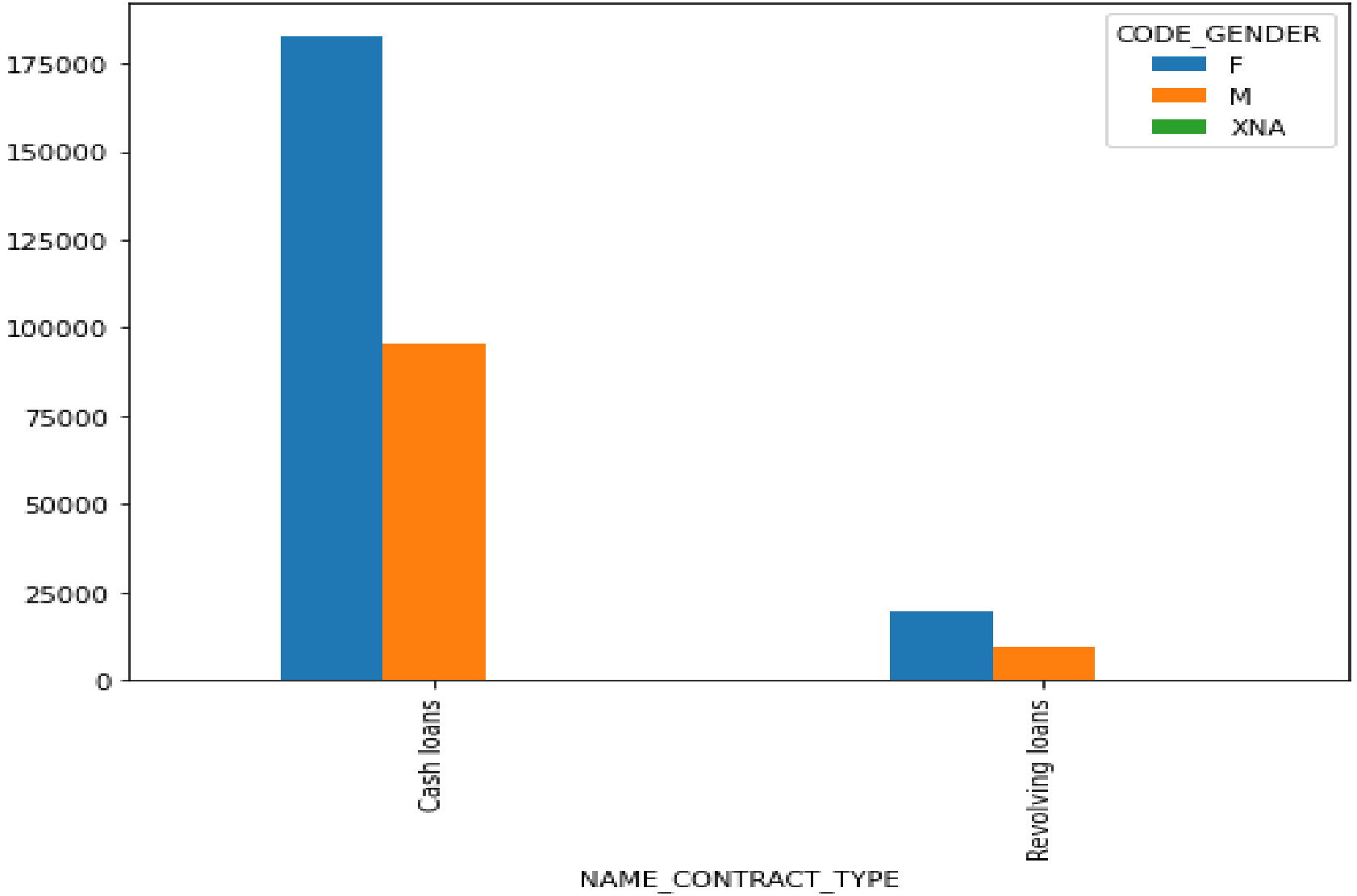
- Income of Office apartment clients is more followed by Municipal Apartment.
- Housing type of with parents has the least income.



CONTRACT TYPE AND GENDER

From the graph we conclude that:

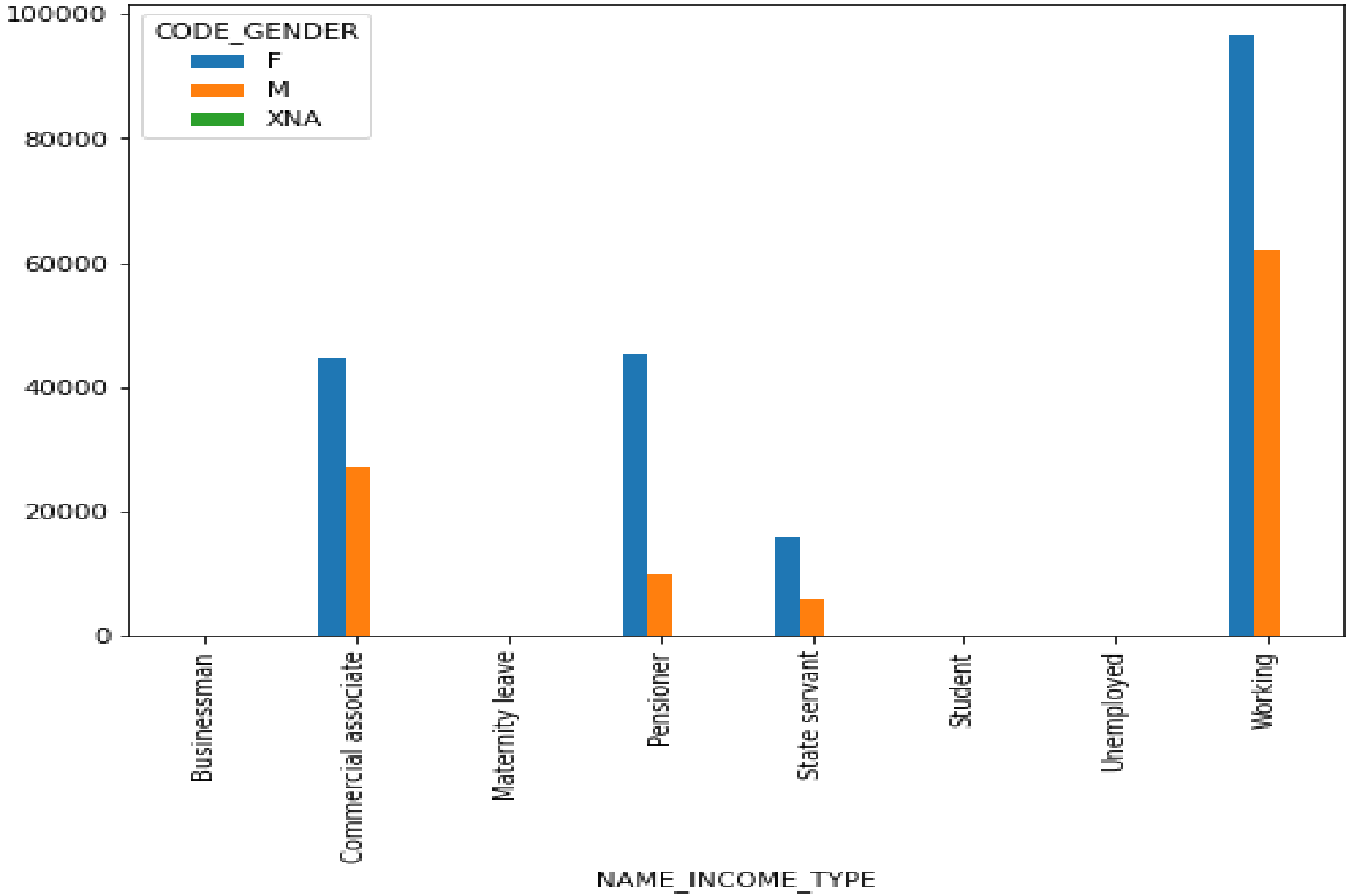
- In general Females tend to take more loans compared to Males
- Female take more Cash Loans as compared to Revolving Loans.
- Male also take more Cash Loans compared to Revolving Loans.



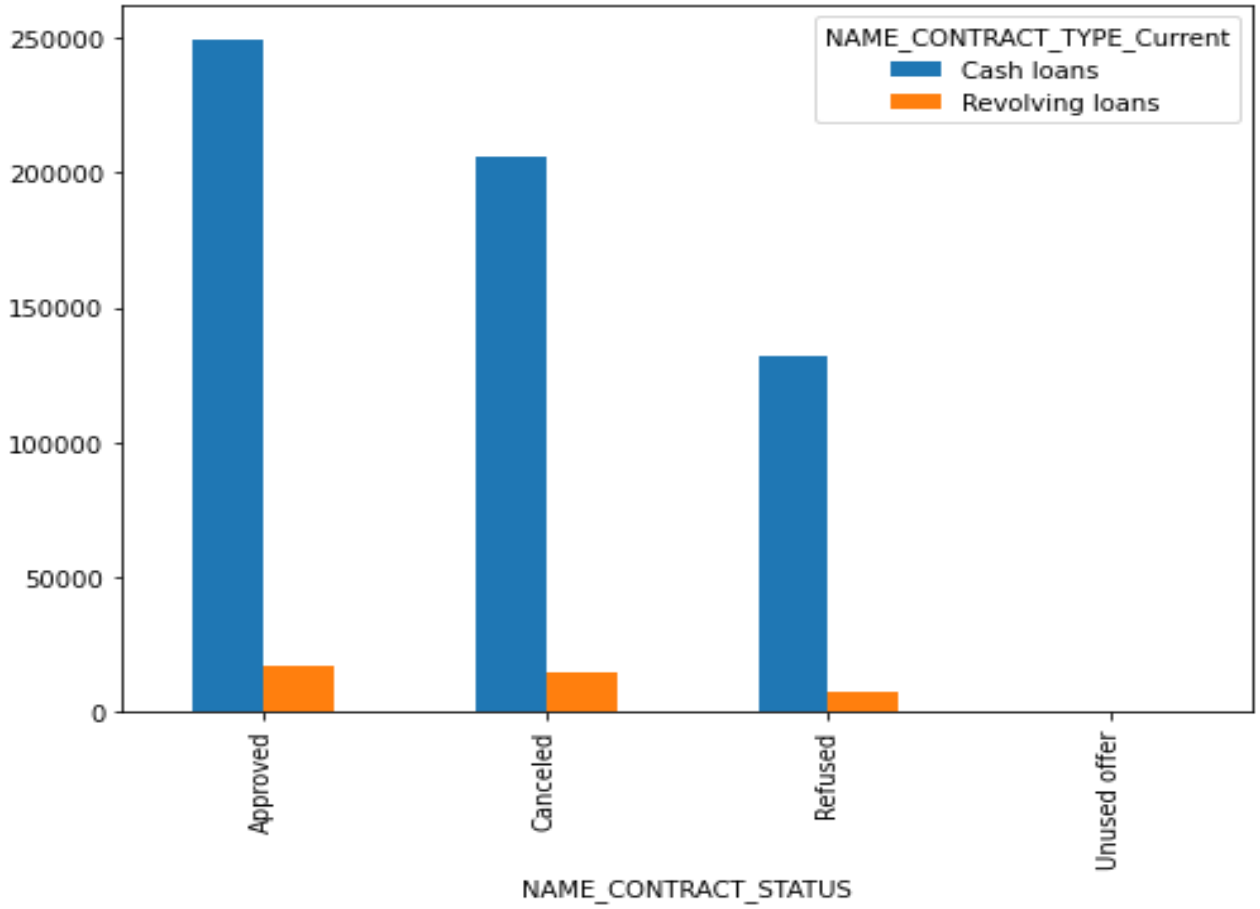
INCOME TYPE AND GENDER

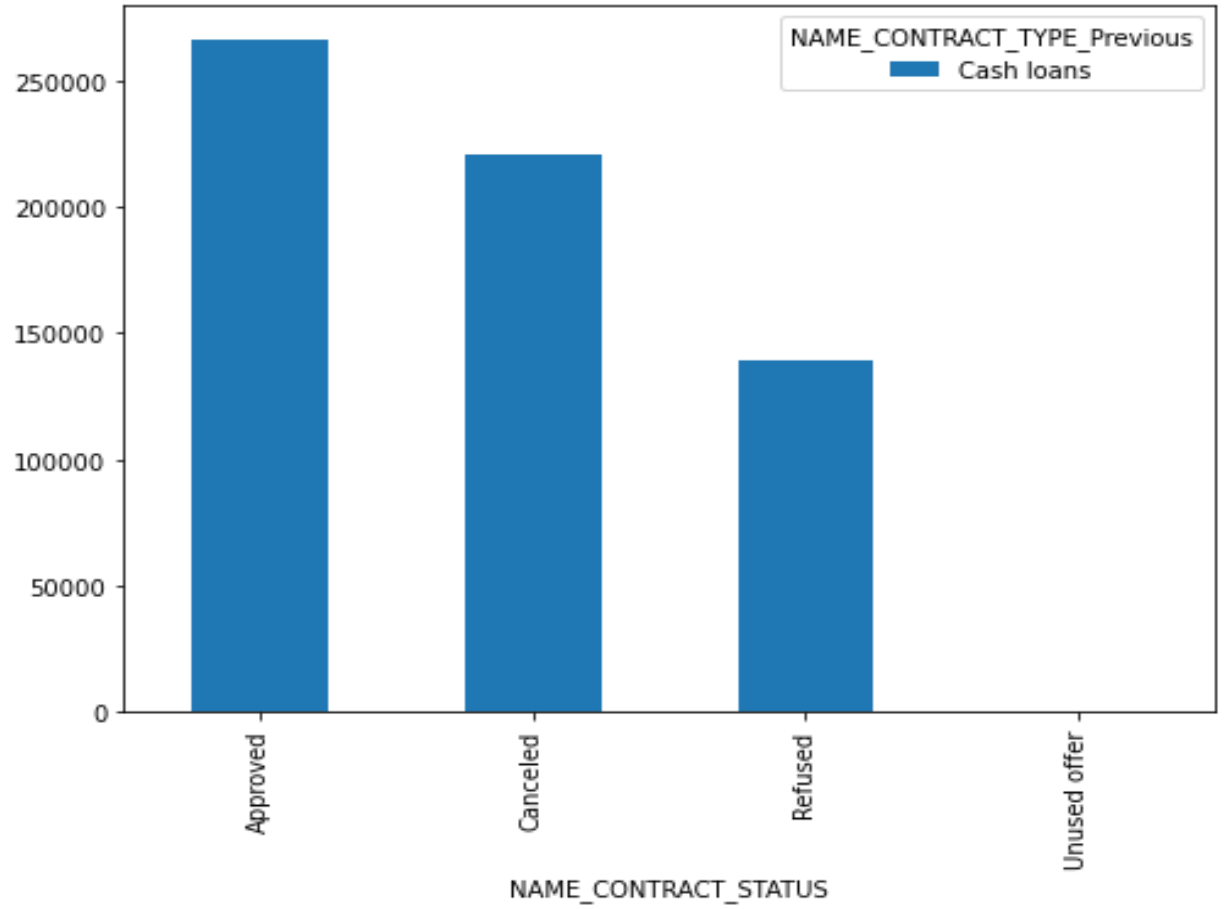
From the graph we conclude that:

- Out of all the income type most income is generated by Female as compared to men.
- Max income from Females is from working type.
- Even max income from Males is also from working type.
- Least income for Males and Females both are from State Servant.



CONTRACT TYPE AND CONTRACT STATUS

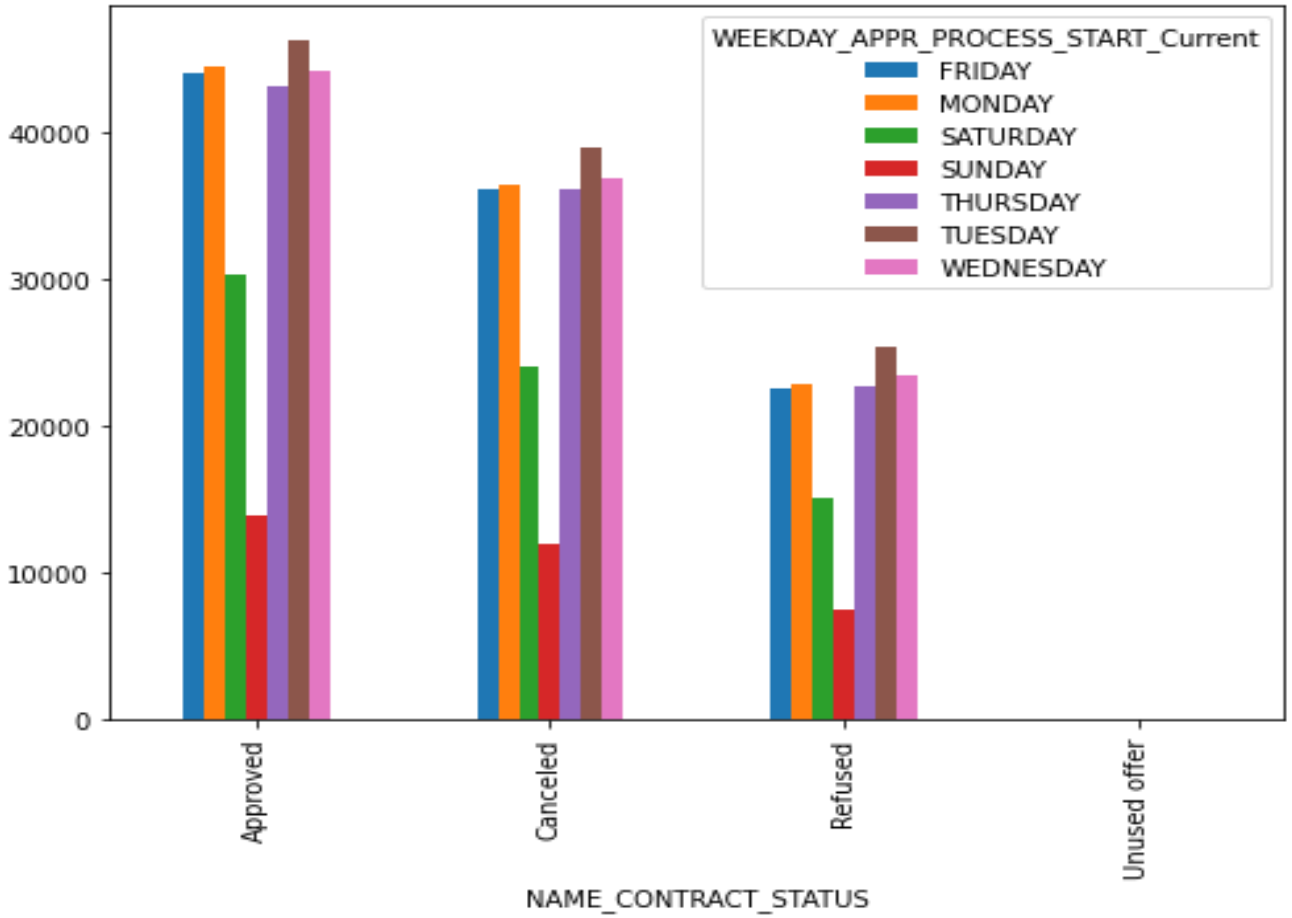


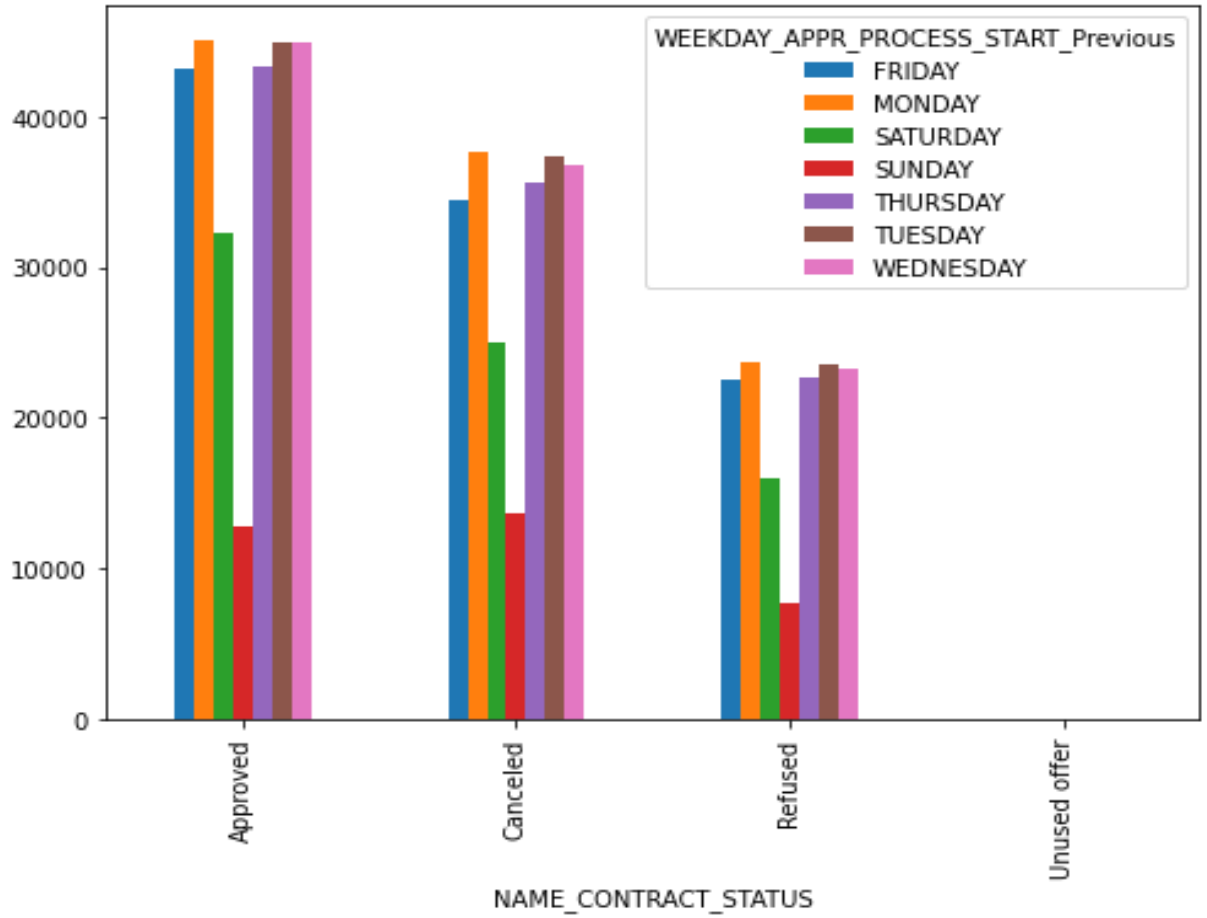


From the graph we conclude that:

- Cash Loans contract type is Approved more.
- It seems Revolving Loans is newly introduced, that is why it is not available in the previous contract type.

WEEKDAY PROCESS AND CONTRACT STATUS





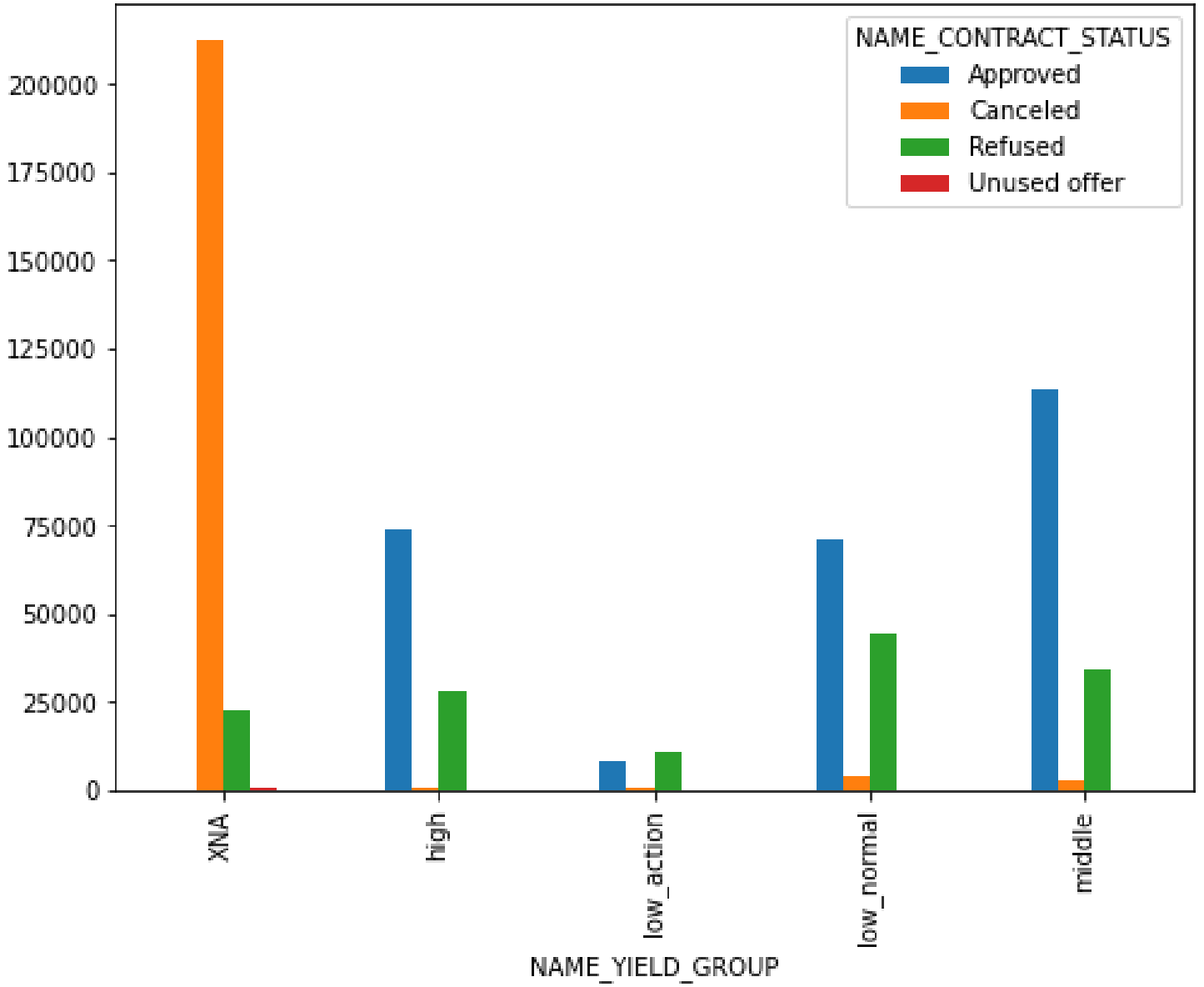
From the graph we conclude that:

- In the current and previous scenarios maximum no. of loans is approved on Monday, Tuesday and Wednesday
- In both the scenarios least no. of approvals is done on Sunday, which can be explained by stating Sunday is a bank holiday.
- Also maximum no. of loans is cancelled and Refused on Monday, Tuesday and Wednesday only.

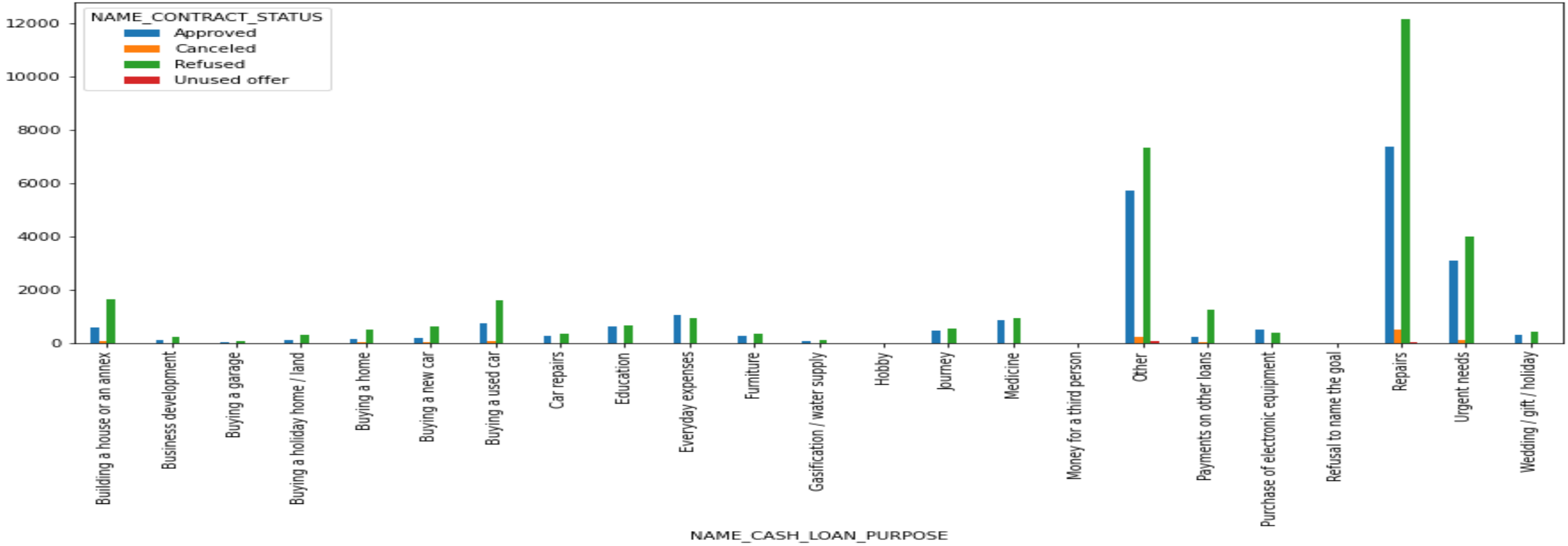
YIELD GROUP AND CONTRACT STATUS

From the graph we conclude that:

- Middle Yield group has highest approval, followed by high and Low_Normal.



CASH LOAN PURPOSE AND CONTRACT STATUS



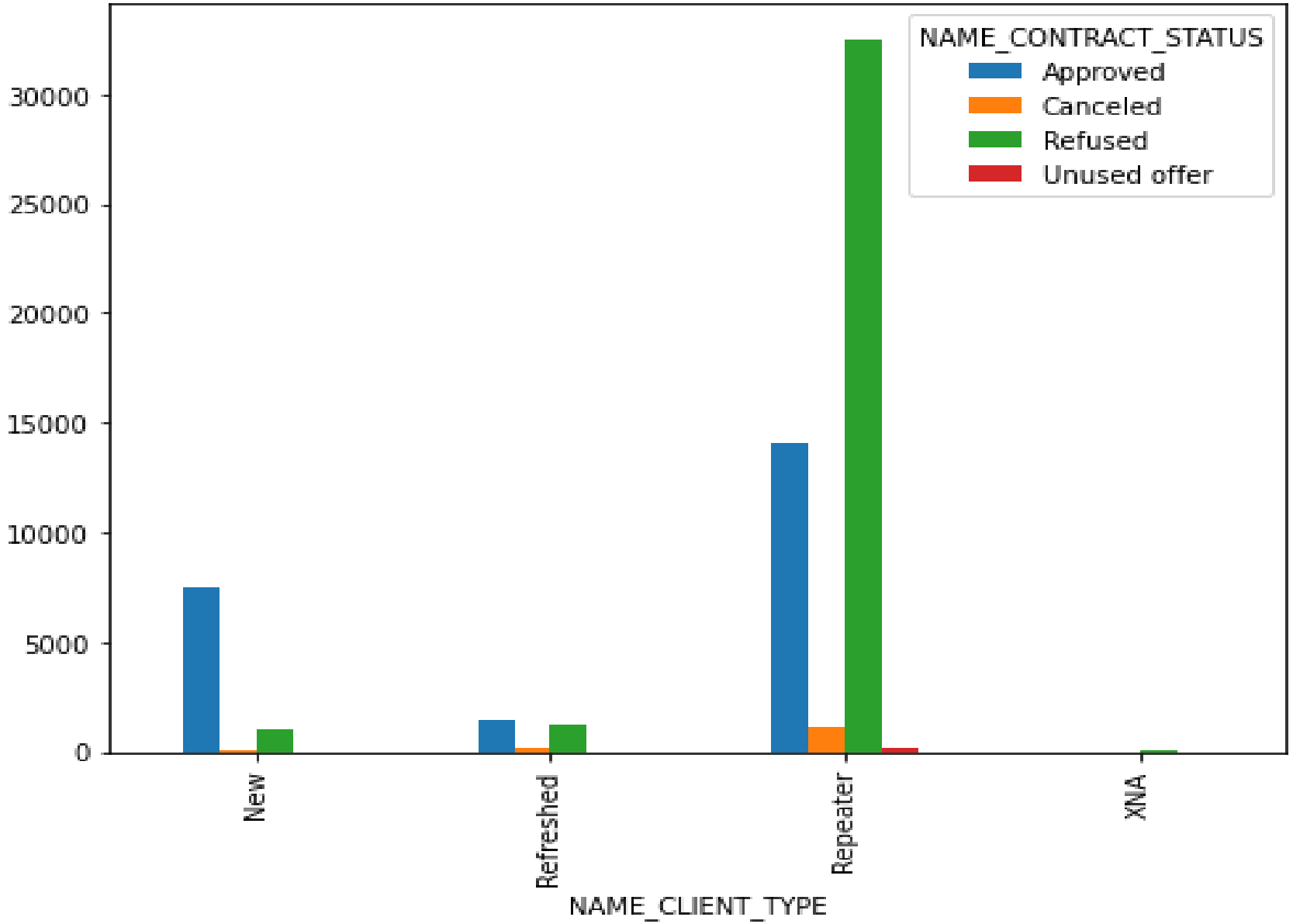
From the graph we conclude that:

- Maximum cash loan was taken for repair purposes.
- At the same time bank has refused loan for repair purposes.

CLIENT TYPE AND CONTRACT STATUS

From the graph we conclude that:

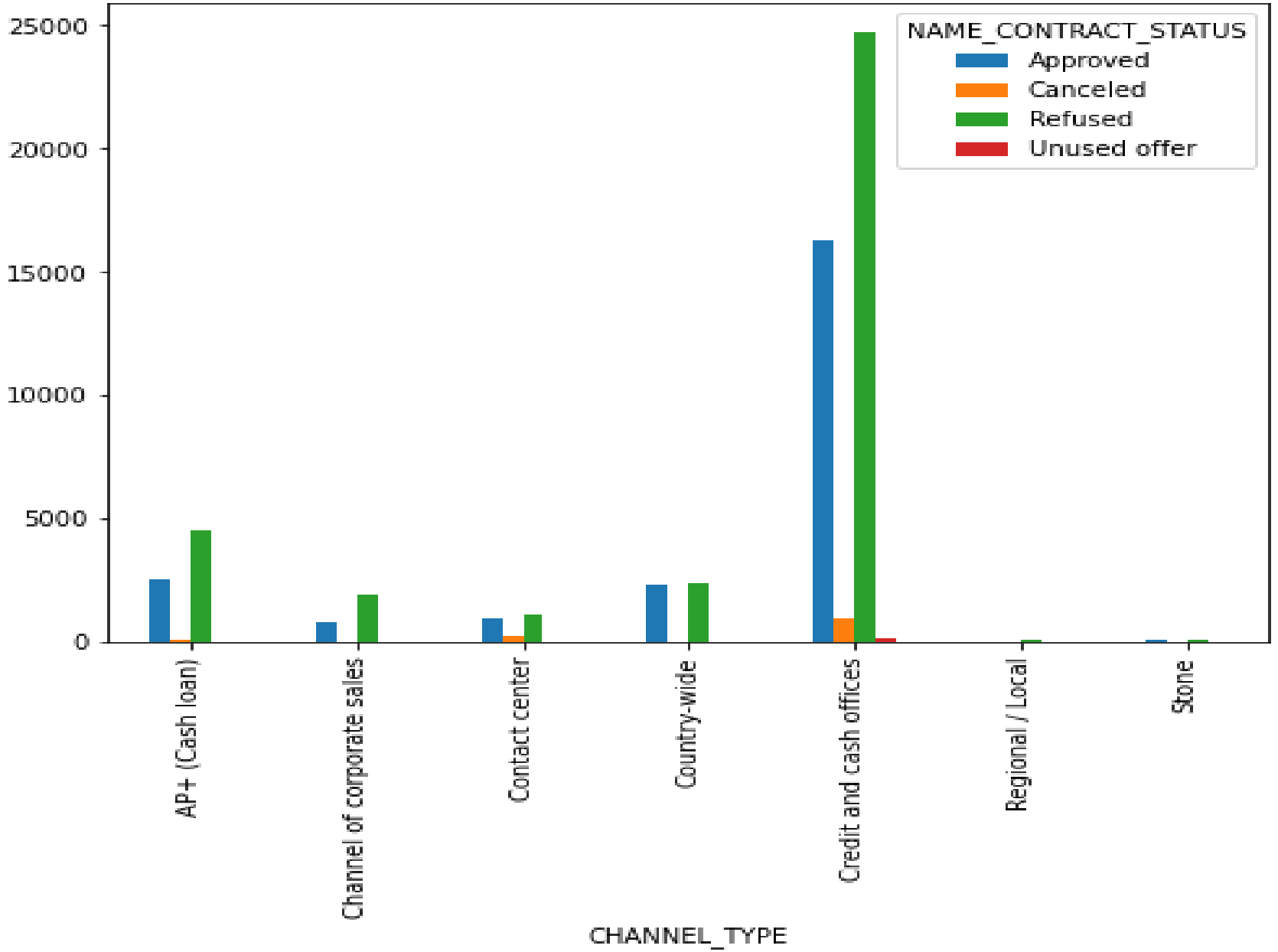
- Repeaters Loan has been approved maximum no. of times.
- At the same time Repeaters has been rejected too.
- New Customer has more approvals compared to Refused.



CHANNEL TYPE AND CONTRACT STATUS

From the graph we conclude that:

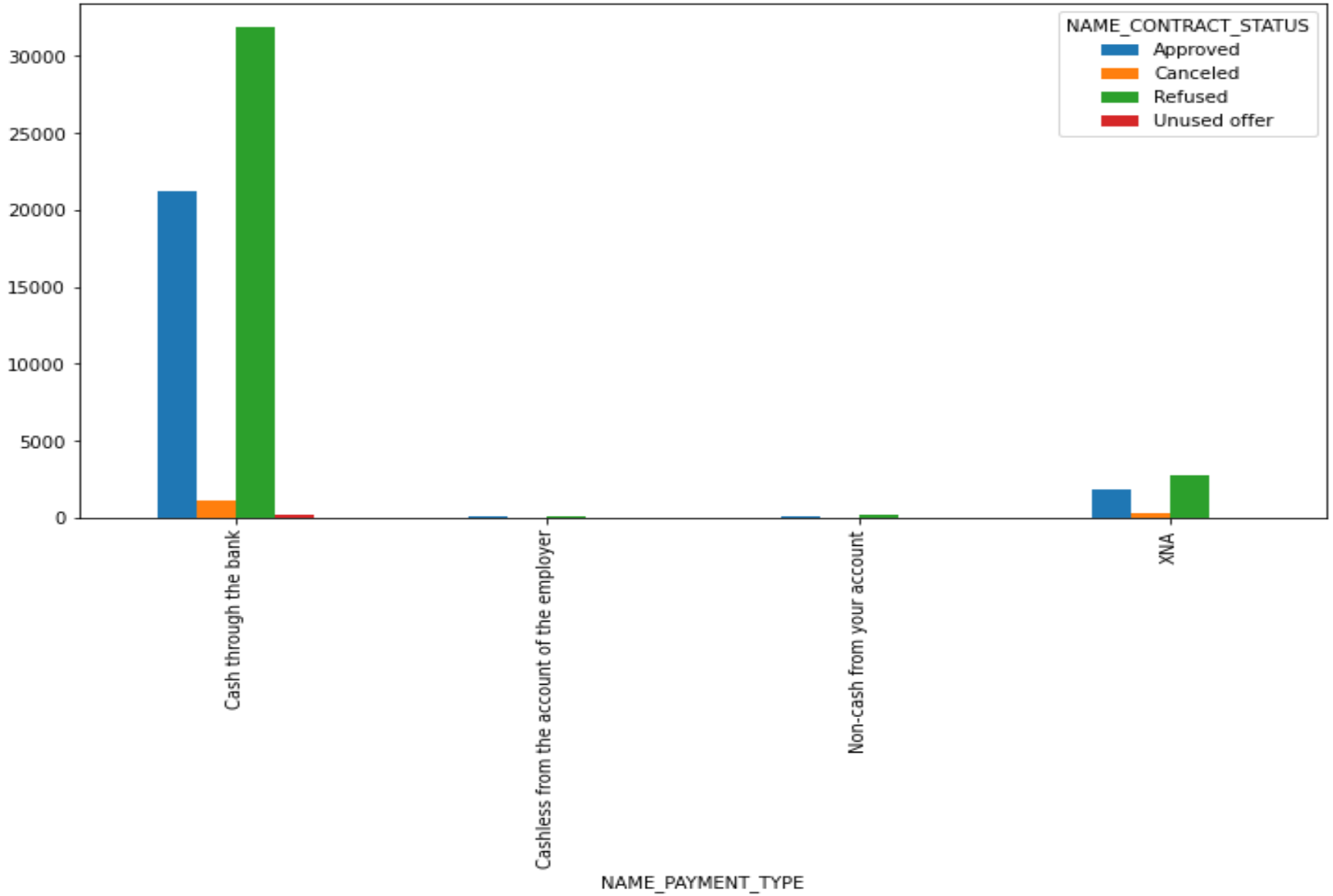
- Channel Credit and cash offices has the maximum no. of enquires.
- As the max enquires from Credit and cash offices approvals and rejection is also from this channel.



PAYMENT TYPE AND CONTRACT STATUS

From the graph we conclude that:

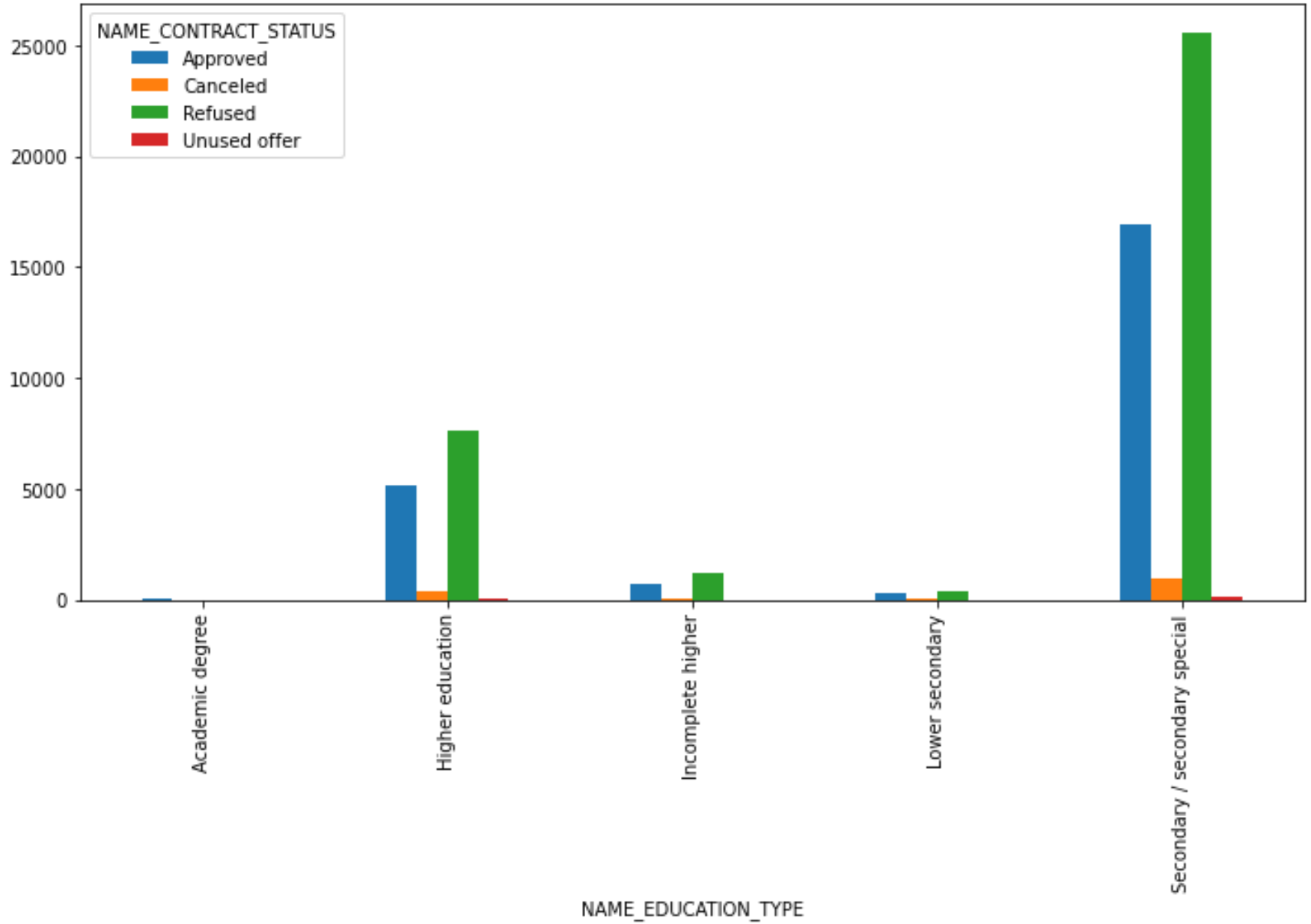
- Bank has contracts paid maximum by Cash.
- Also at the same time they are refused maximum no. of times.



EDUCATION TYPE AND CONTRACT STATUS

From the graph we conclude that:

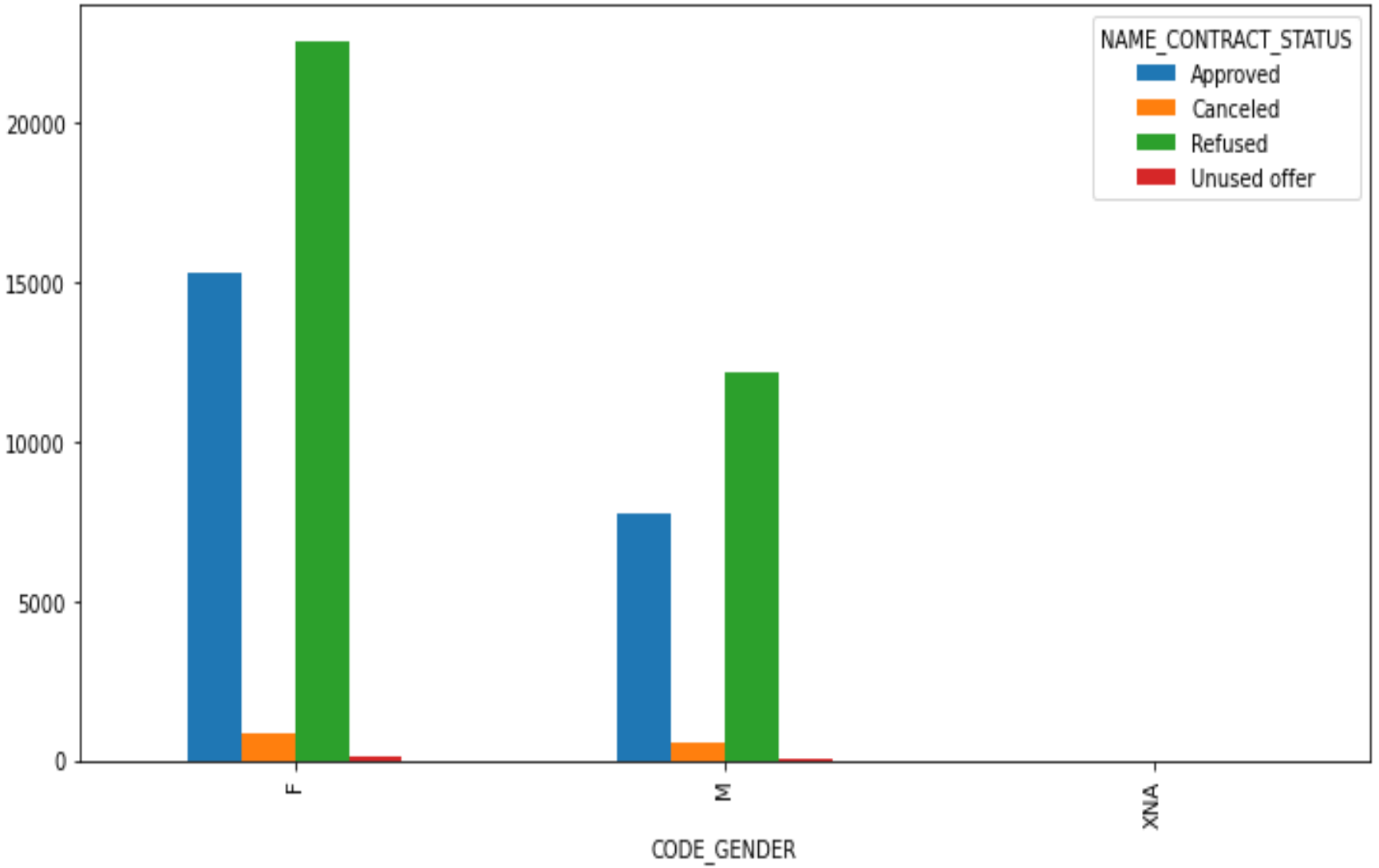
- Secondary Education has maximum approvals and maximum no. of rejection.
- Followed by Higher Education.



GENDER AND CONTRACT STATUS

From the graph we conclude that:

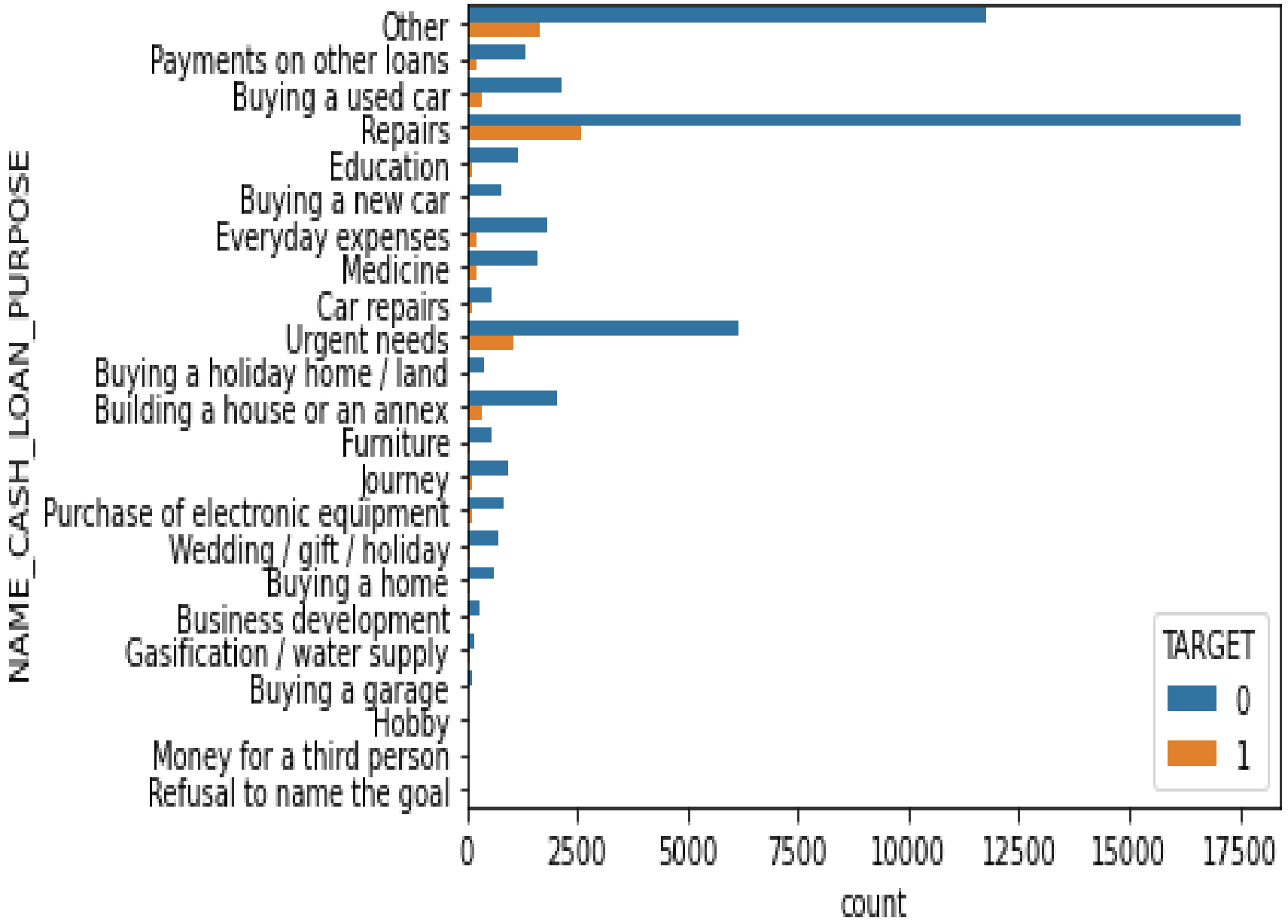
- Female has the maximum no. of loan approvals as compared to Male.
- Also more loans were refused to Females as compared to Males.



CASH LOAN PURPOSE AND TARGET

From the graph we conclude that:

- Repairs have the most no. of customers with No Payment Difficulties.



CONCLUSION

Below is the conclusion for the Analysis:

- ✓ Bank already has a big amount of Cash Loans customer, and less amount of Revolving Customers, so bank should focus more on Revolving Loan customers.
- ✓ Though the Income of Males is more but Females tend to take more loans, bank should focus more on Males.
- ✓ Businessman have the highest credit amount but they tend to take the least amount of loans, banks should be more focused on Businessman as it bring huge revenue.
- ✓ Banks does its maximum business from Monday to Wednesday from 9AM to 6PM i.e. normal working hours.
- ✓ Bank should consider giving loans to applicants who stays with parents as they give less payment difficulties to bank.
- ✓ Bank should focus on Single/Not Married as they are less likely to default on Loans.
- ✓ Secondary Education take more loans but at the same time they even cause payment difficulties to bank, rather bank should focus on Higher Education as they are less likely to cause payment difficulties compared to Secondary Education.
- ✓ For successful payments banks should focus more on contract type Student, pensioner and Businessman with housing type other than Co-op apartment.
- ✓ Banks in cash loan purpose should focus on Repairs followed by Others and Urgent Needs as they are less likely to default on loans.

FACTORS TO BE CONSIDERED BEFORE APPROVING LOANS.

- Income of Customers.
- Credit Amount
- Loan Purpose
- Education
- Income type
- Annuity Amount
- Housing Type
- Family Status
- Previous Loan Status