Purpose of Credit EDA Case Study:

- Credit risk analysis performed by a company/bank/financial institution to determine whether the client/applicant is eligible for loan approval or not.

- Clients who present a difficulty paying their installment are charged a high interest rate or can get the loan rejection or loan amount could get reduced.

- This will ensure that the consumers capable of repaying the loan are not rejected.

Exploratory Data Analysis Steps:

1. Data sourcing and understanding the data.

2. Checking for data imbalance.

3. Performing Data cleaning which involves removing irregularities in data by treating missing values and outlier handling.

4. Variable transformation and creation.

5. Fixing invalid values and filter the data.

6. Performing Univariate and Bivariate analysis.

7. Merging the application data with previous application data.

8. Performing Univariate and Bivariate analysis on merged data.

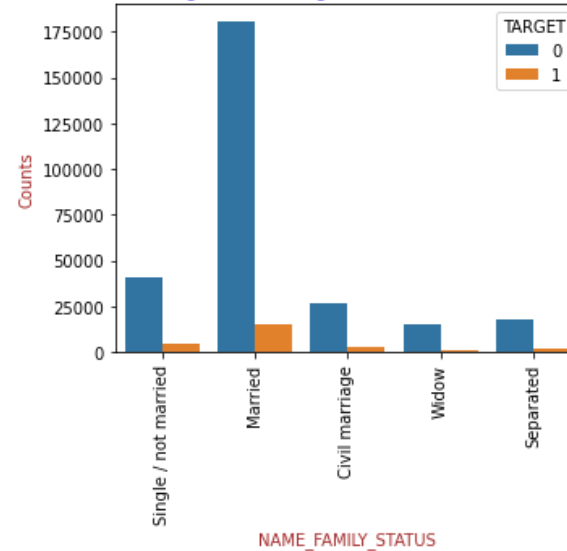9. Insights and recommendations for the company/bank.

1st plot - Univariate analysis using Pie chart
2nd and 3rd plots – Segmented Univariate analysis on **Family Status and Target variable.**

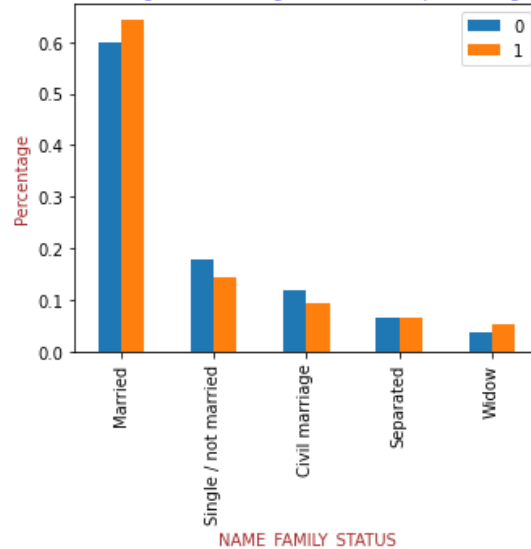Percentage of each category in NAME_FAMILY_STATUS

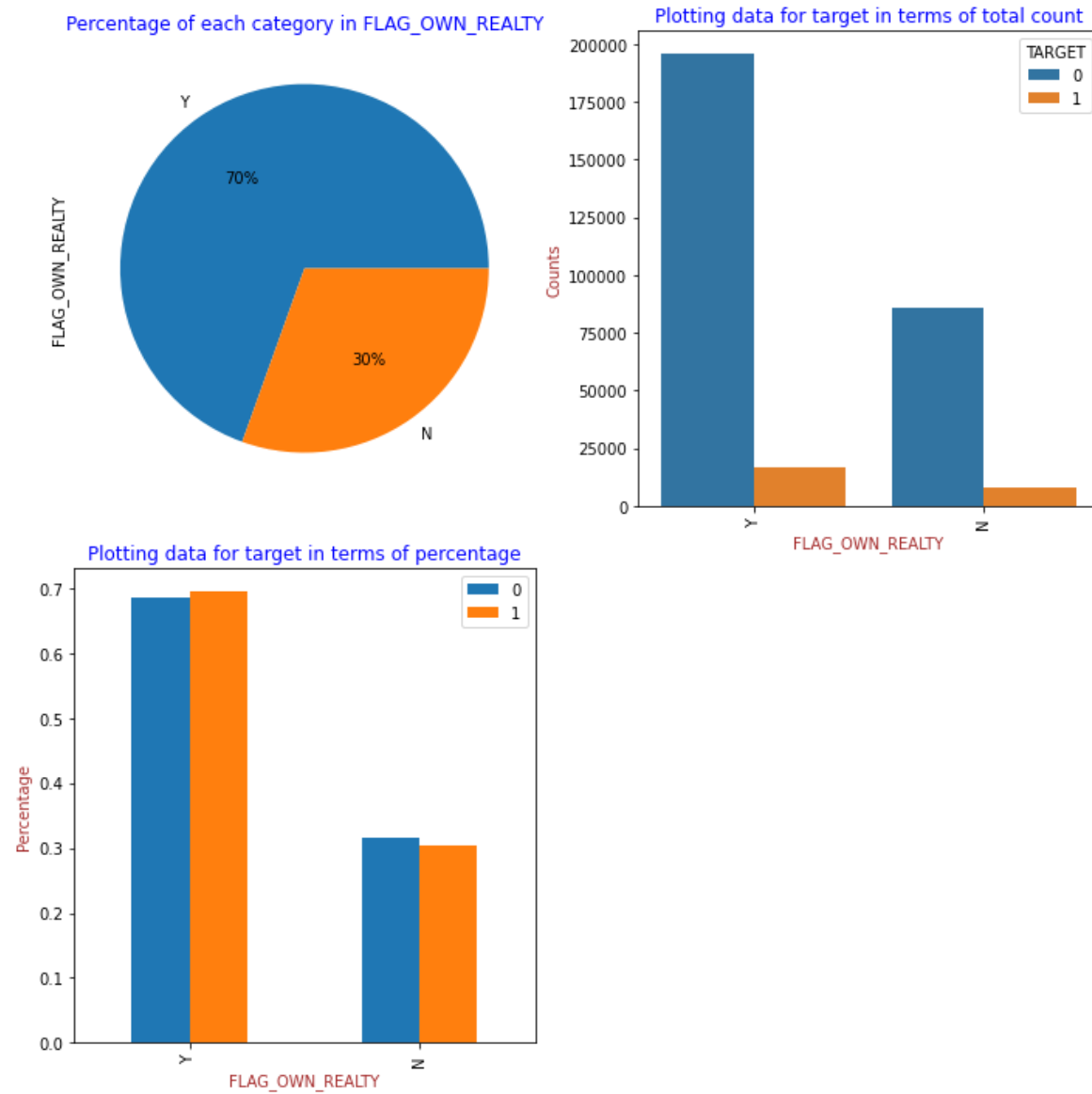Plotting data for target in terms of total count
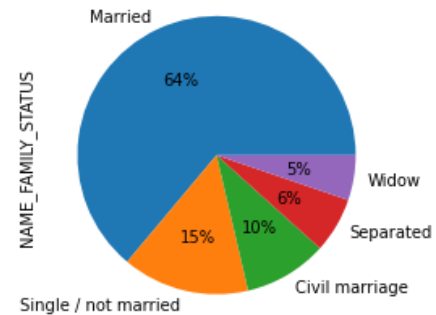
Plotting data for target in terms of percentage

Among the family status , we can see that Married and widowed clients are likely to default when compared with others.

1ˢᵗ plot - Univariate analysis using Pie chart
2ⁿᵈ and 3ʳᵈ plots - Segmented Univariate analysis on **House/Flat owner and Target variable.**
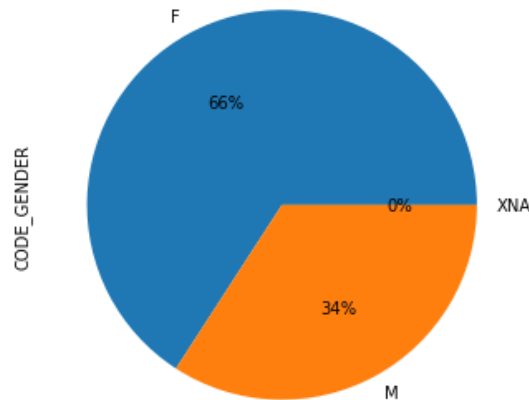

Count of clients who own realty are more than clients who don't. However we can see that the later ones are weak candidates to default.
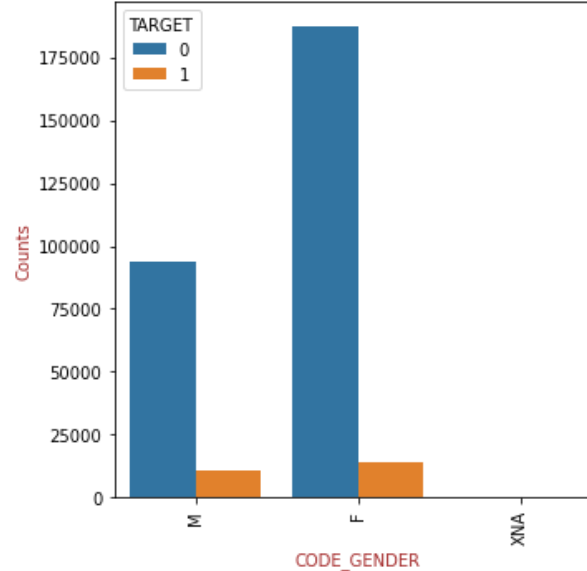
1st plot - Univariate analysis using Pie chart
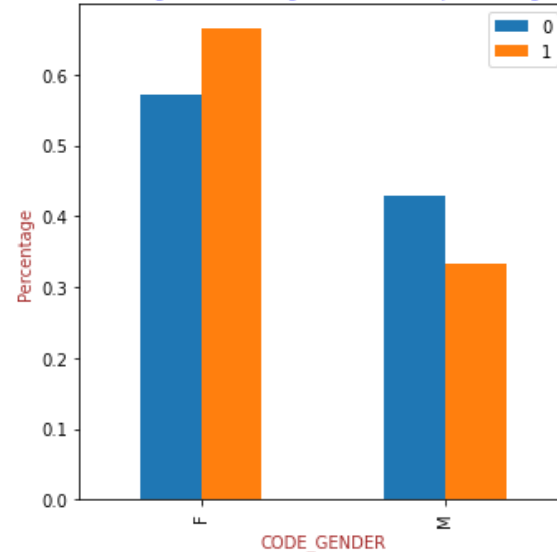2nd and 3rd plots - Segmented Univariate analysis on **Gender and Target variable.**
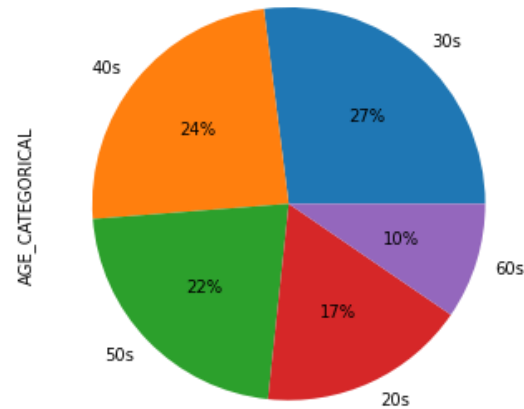

More number female clients have applied for loans than males. But also chances to default the loan is higher in females.
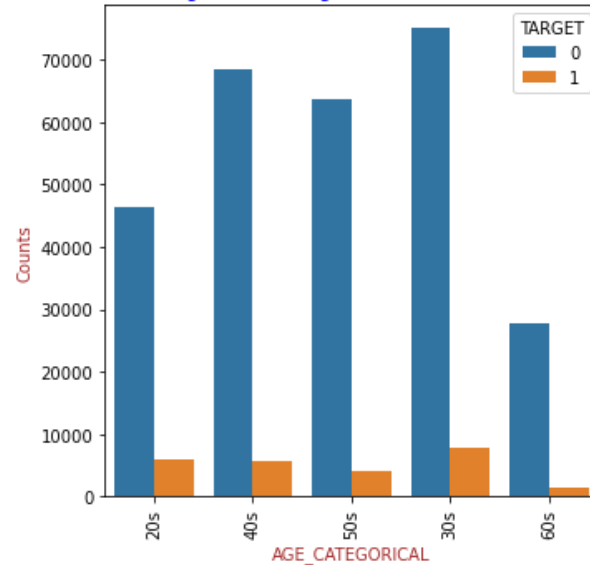
1st plot - Univariate analysis using Pie chart
2nd and 3rd plots - Segmented Univariate analysis on **Age Category and Target variable.**
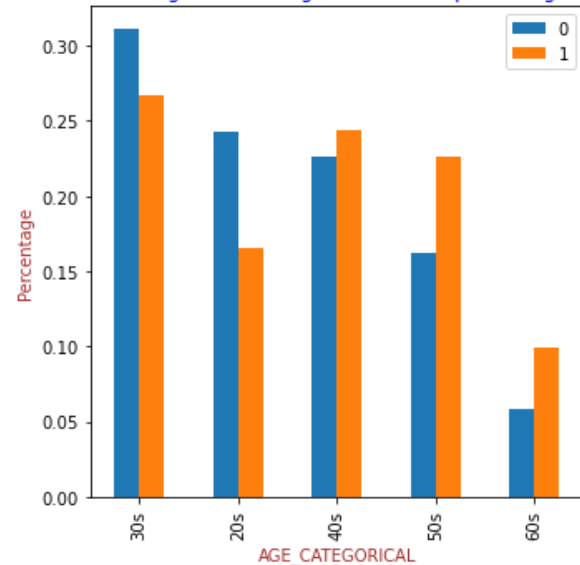

Percentage of each category in AGE_CATEGORICAL

Plotting data for target in terms of total count

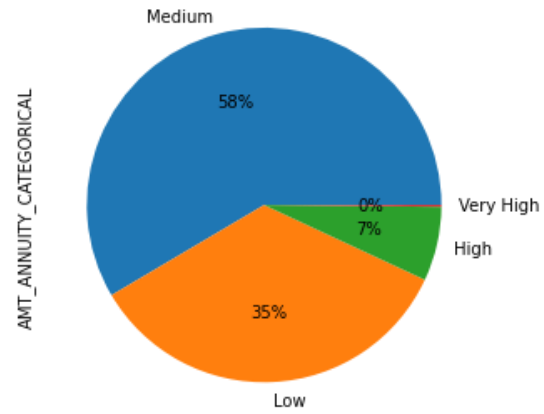Plotting data for target in terms of percentage

Clients having age less than 40 have lower rate of default compared to the other age group.

1ˢᵗ plot - Univariate analysis using Pie chart
2ⁿᵈ and 3ʳᵈ plots - Segmented Univariate analysis on **Annuity amount and Target variable.**

Count of clients falling under medium annuity range is higher than High and Low annuity range. But the defaulters are seen to be more in High and Low annuity range.
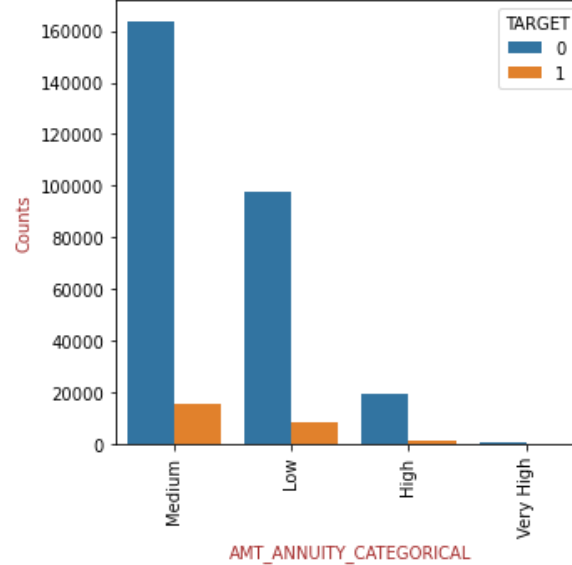
1st plot - Univariate analysis using Pie chart
2nd and 3rd plots - Segmented Univariate analysis on **Education and Target variable.**
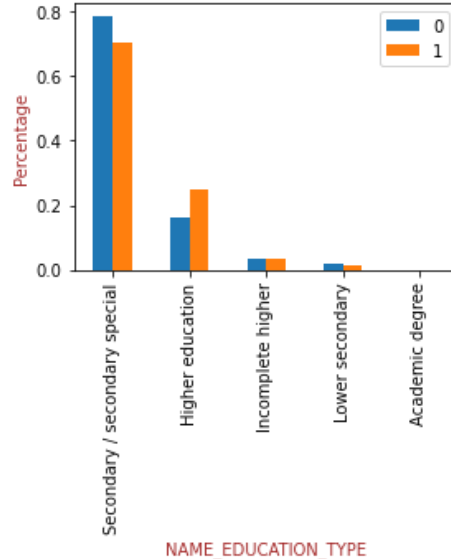
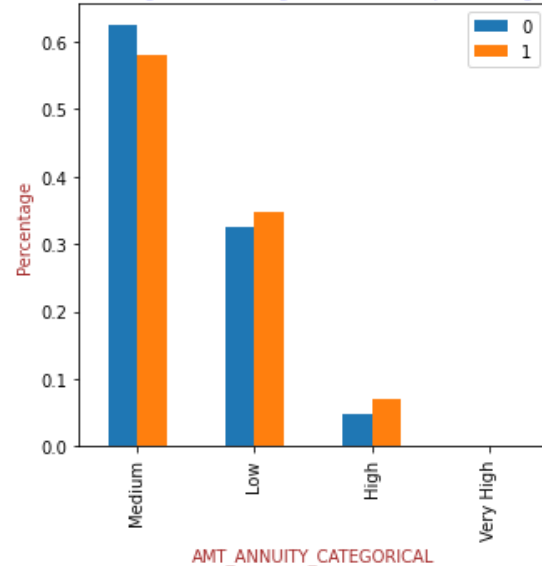Clients pursuing Secondary/Secondary special education have high chances to default compared to Education types.

1ˢᵗ plot - Univariate analysis using Pie chart
2ⁿᵈ and 3ʳᵈ plots - Segmented Univariate analysis on **Housing type and Target variable.**
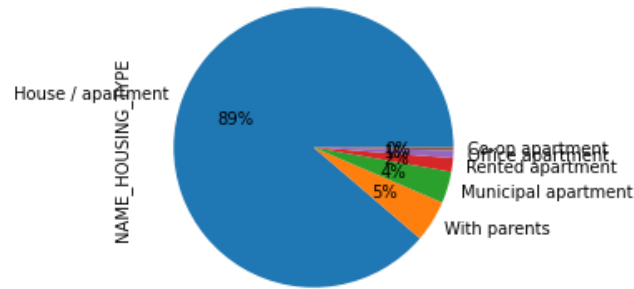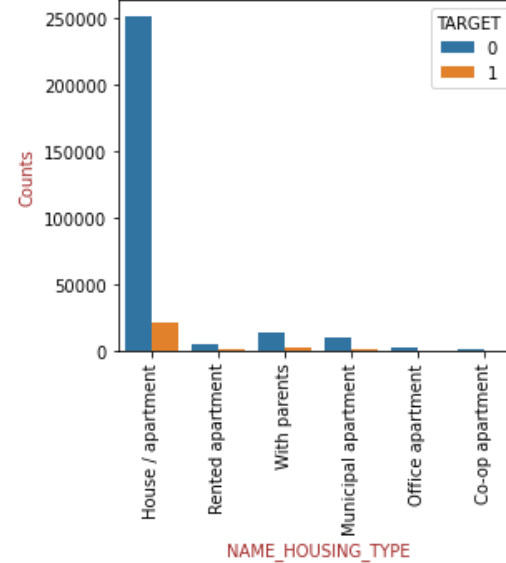
In housing type, clients having House/apartments are likely to default the loans.
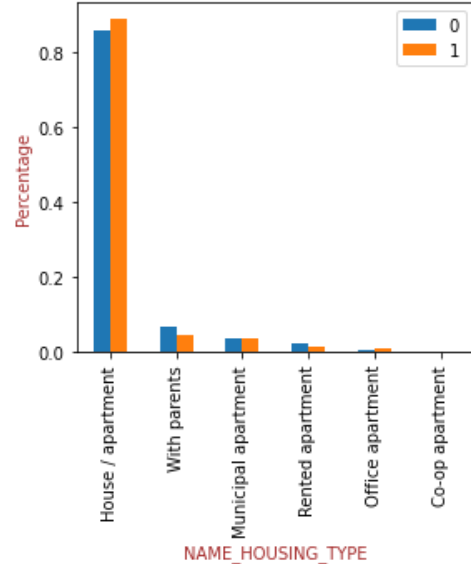
Segmented Univariate analysis on **Organization type and Target variable.**
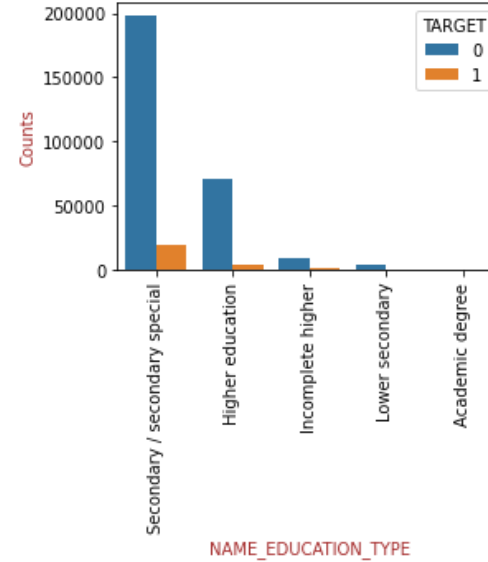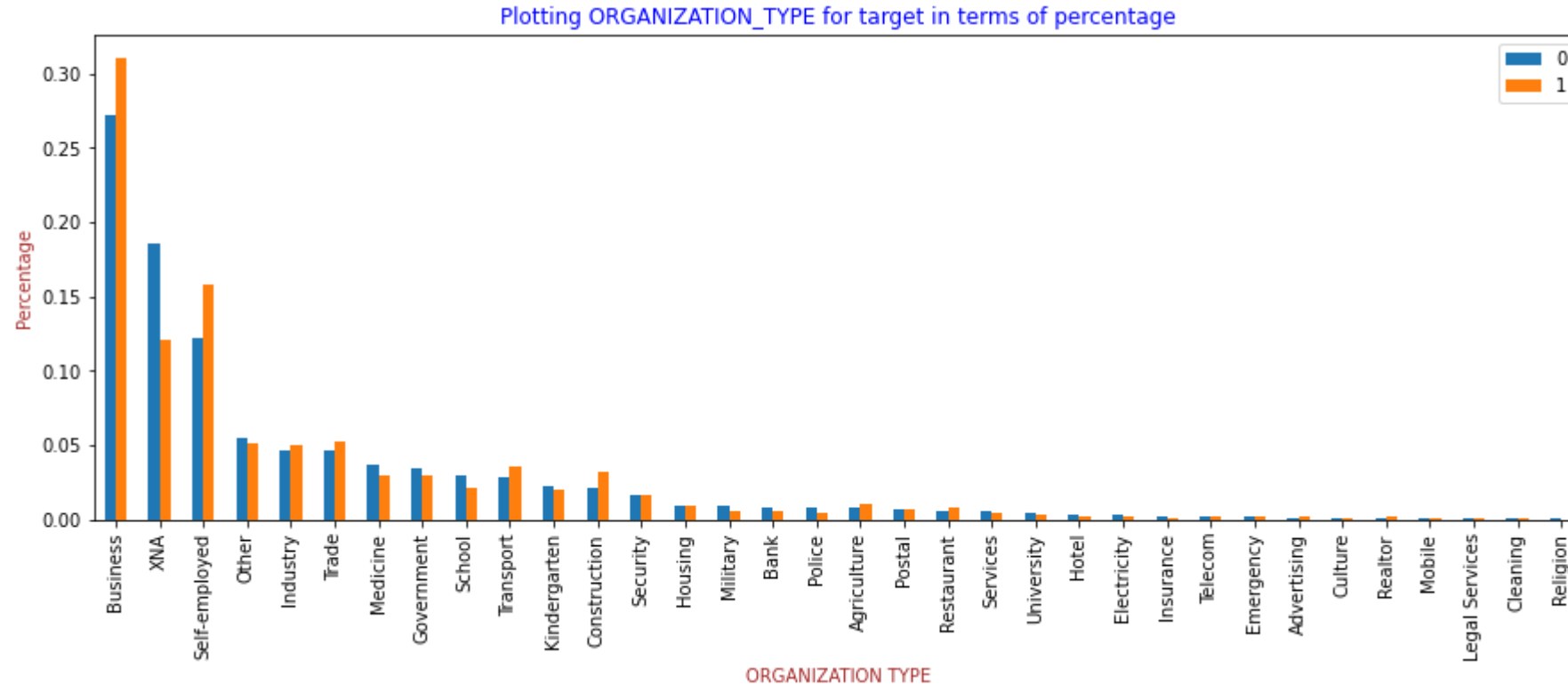

Self-employed clients and the one who own businesses tend to be in defaulters book.

Bivariate analysis on **Goods price amount and Credit amount.**

AMT_GOODS_PRICE Vs AMT_CREDIT

Credit amount increases with increase in goods price amounts.

Multivariate analysis on **Organization type and Target variable.**

Age vs Income of each Occupation Type

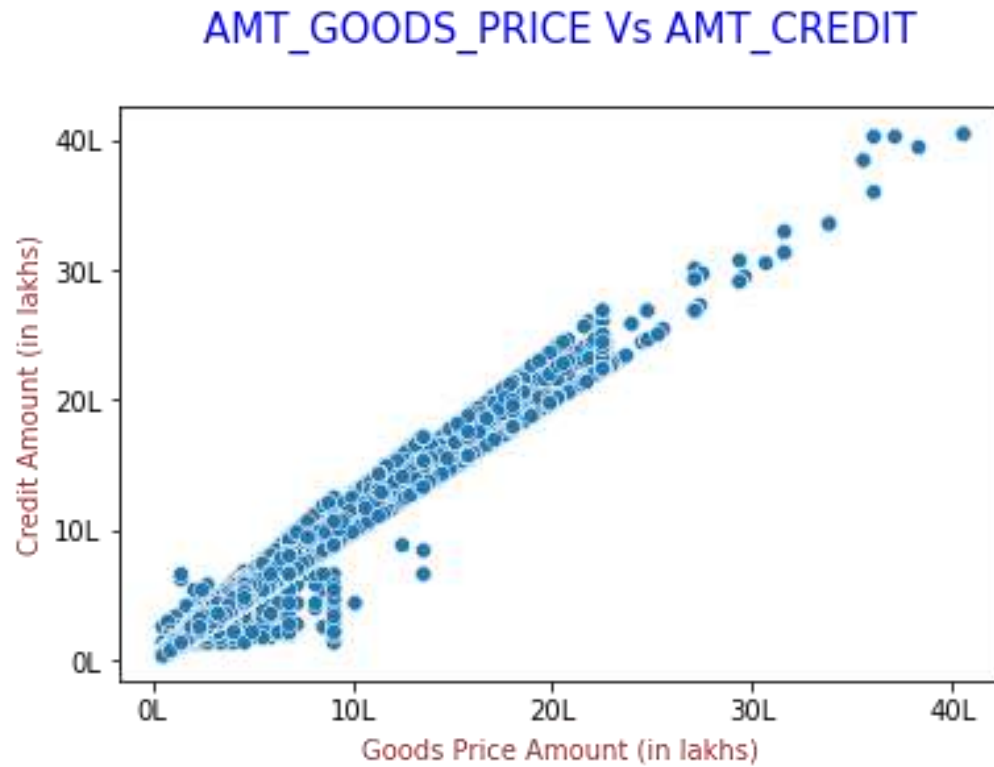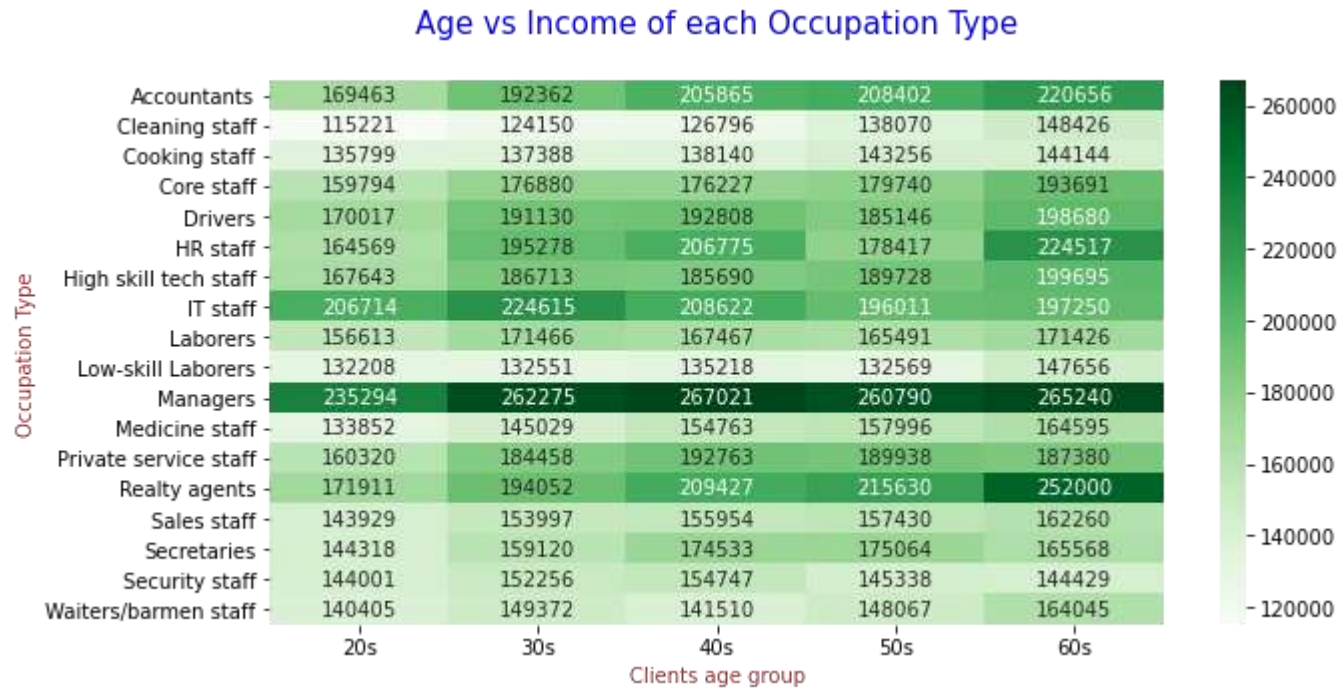| Occupation Type	20s	30s	40s	50s	60s
Accountants	169463	192362	205865	208402	220656
Cleaning staff	115221	124150	126796	138070	148426
Cooking staff	135799	137388	138140	143256	144144
Core staff	159794	176880	176227	179740	193691
Drivers	170017	191130	192808	185146	198680
HR staff	164569	195278	206775	178417	224517
High skill tech staff	167643	186713	185690	189728	199695
IT staff	206714	224615	208622	196011	197250
Laborers	156613	171466	167467	165491	171426
Low-skill Laborers	132208	132551	135218	132569	147656
Managers	235294	262275	267021	260790	265240
Medicine staff	133852	145029	154763	157996	164595
Private service staff	160320	184458	192763	189938	187380
Realty agents	171911	194052	209427	215630	252000
Sales staff	143929	153997	155954	157430	162260
Secretaries	144318	159120	174533	175064	165568
Security staff	144001	152256	154747	145338	144429
Waiters/barmen staff	140405	149372	141510	148067	164045

Clients age group

Self-employed clients and the one who own businesses tend to be in defaulters book.
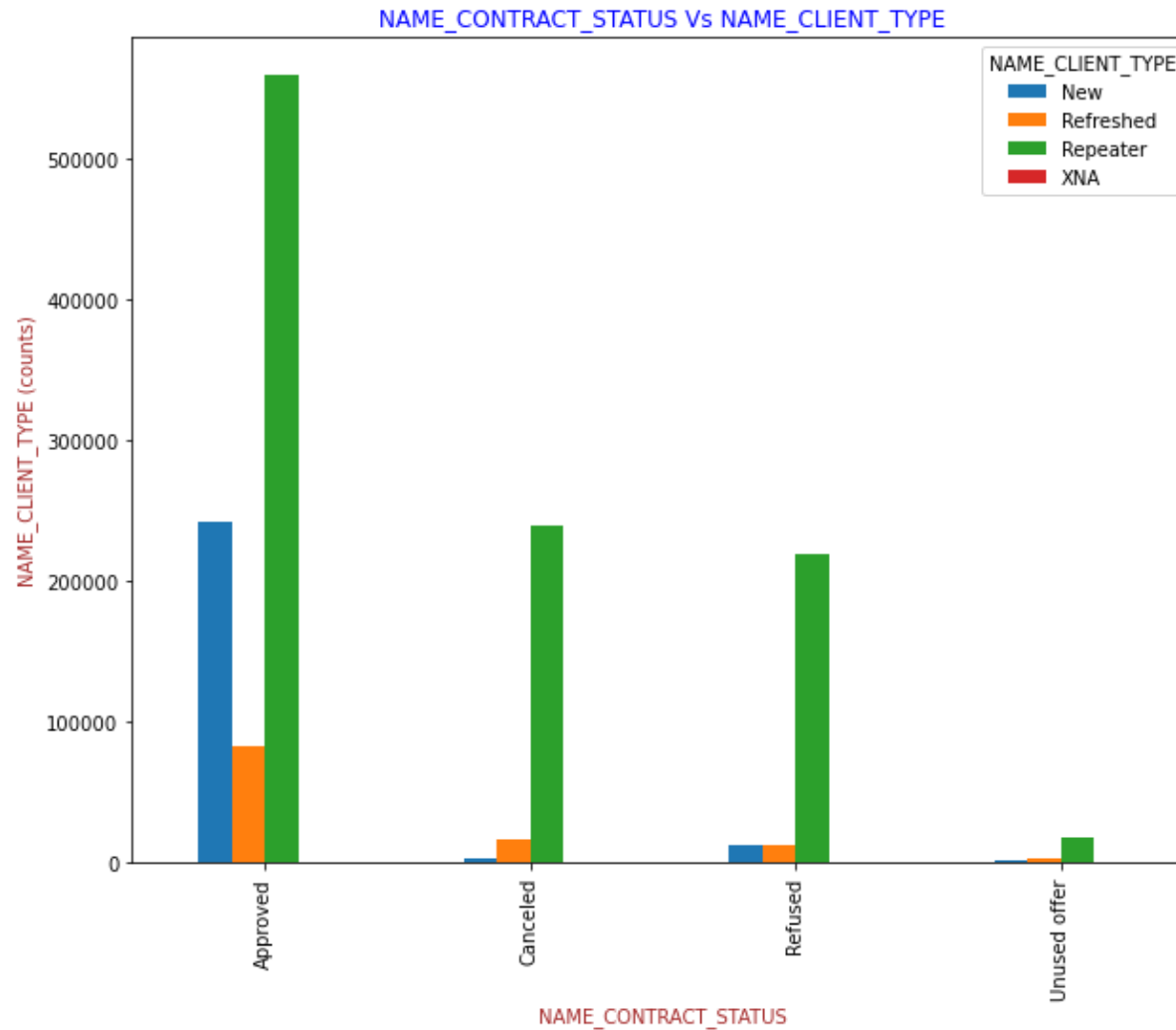
Bivariate analysis on **merged data** – Application data and previous application data.
Contract Type and Contract Status.
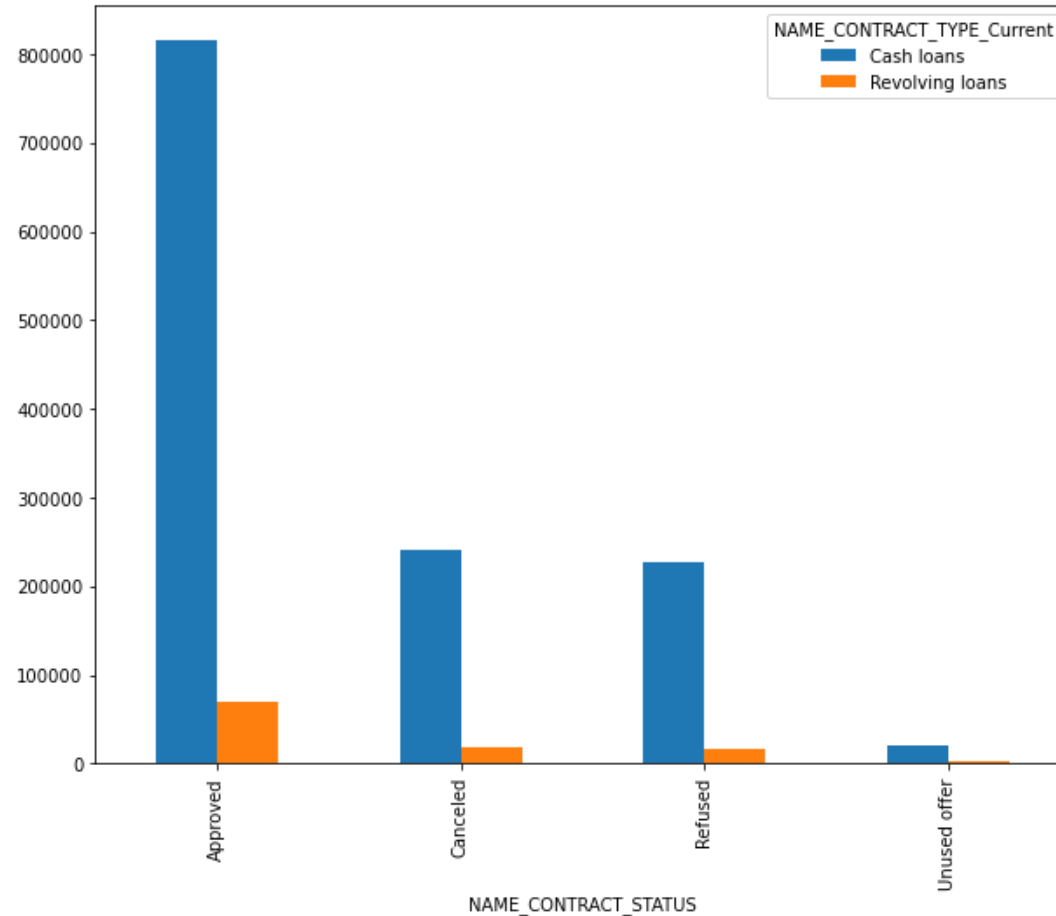
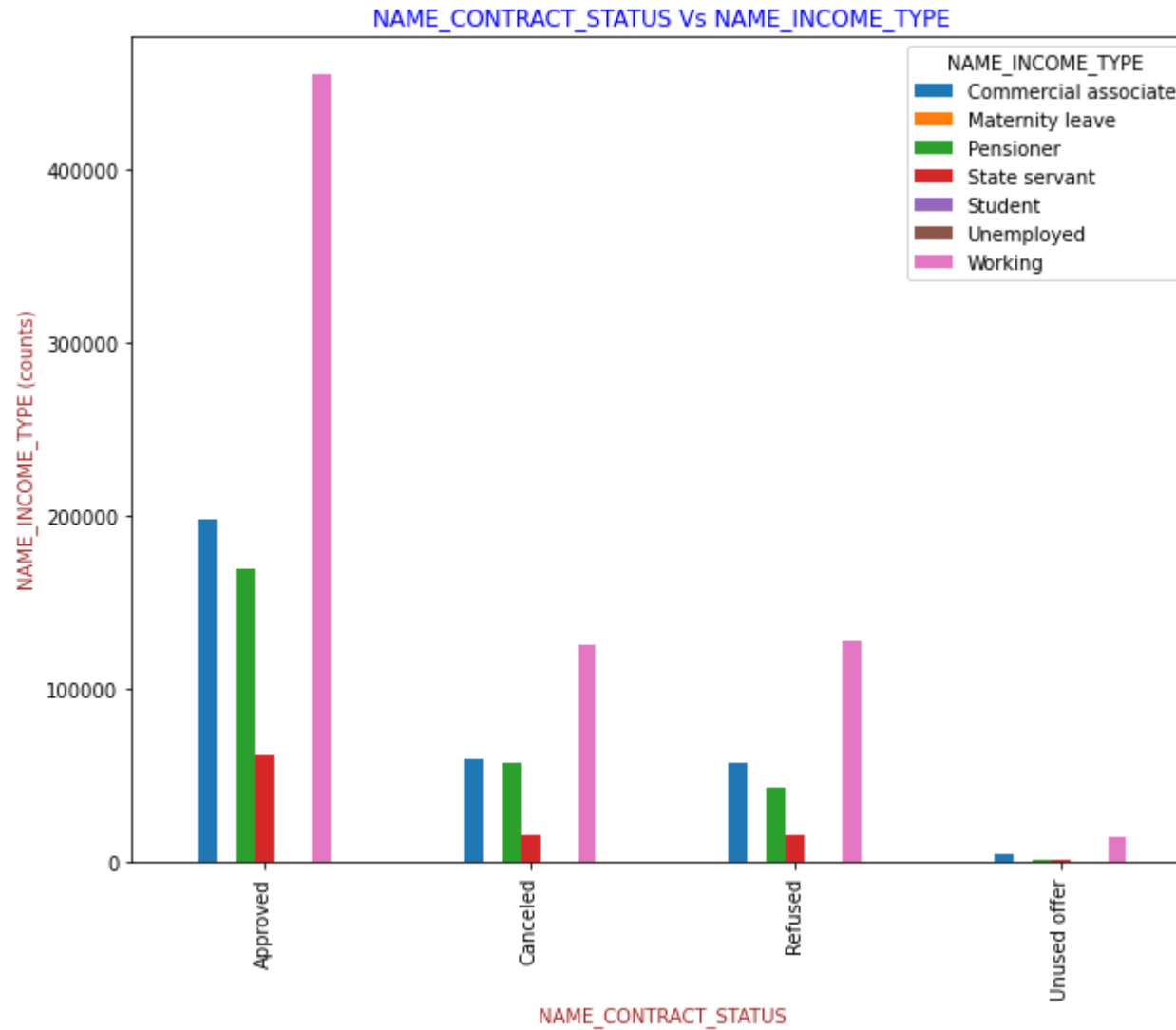Cash loans have highest number of Approved loans.

Bivariate analysis on **merged data** – Application data and previous application data.
Client Type and Contract Status.

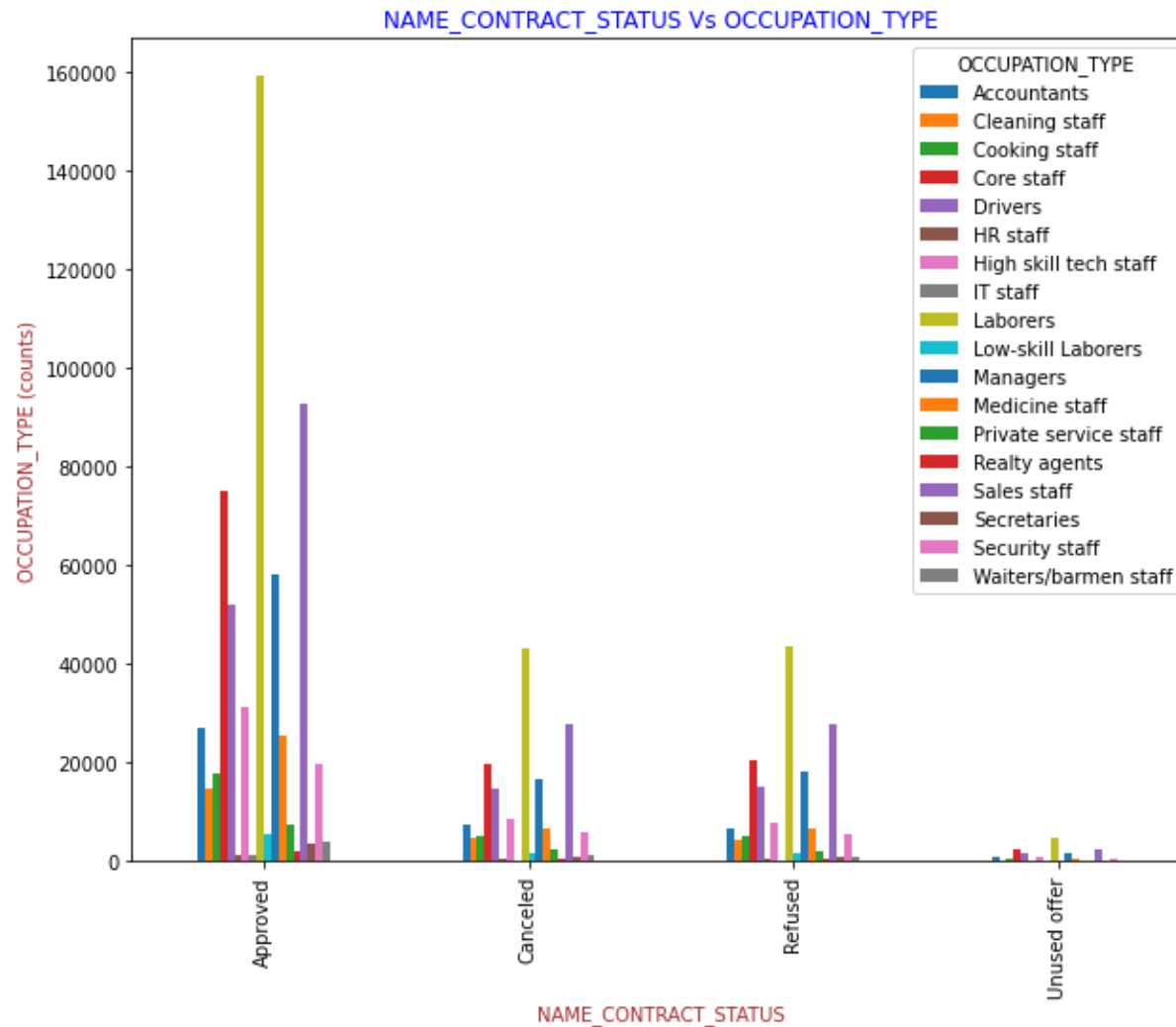Clients, who had applied repeatedly, have highest number of approvals.

Bivariate analysis on **merged data** – Application data and previous application data.
Income Type and Contract Status.

Highest number of approvals for working clients.
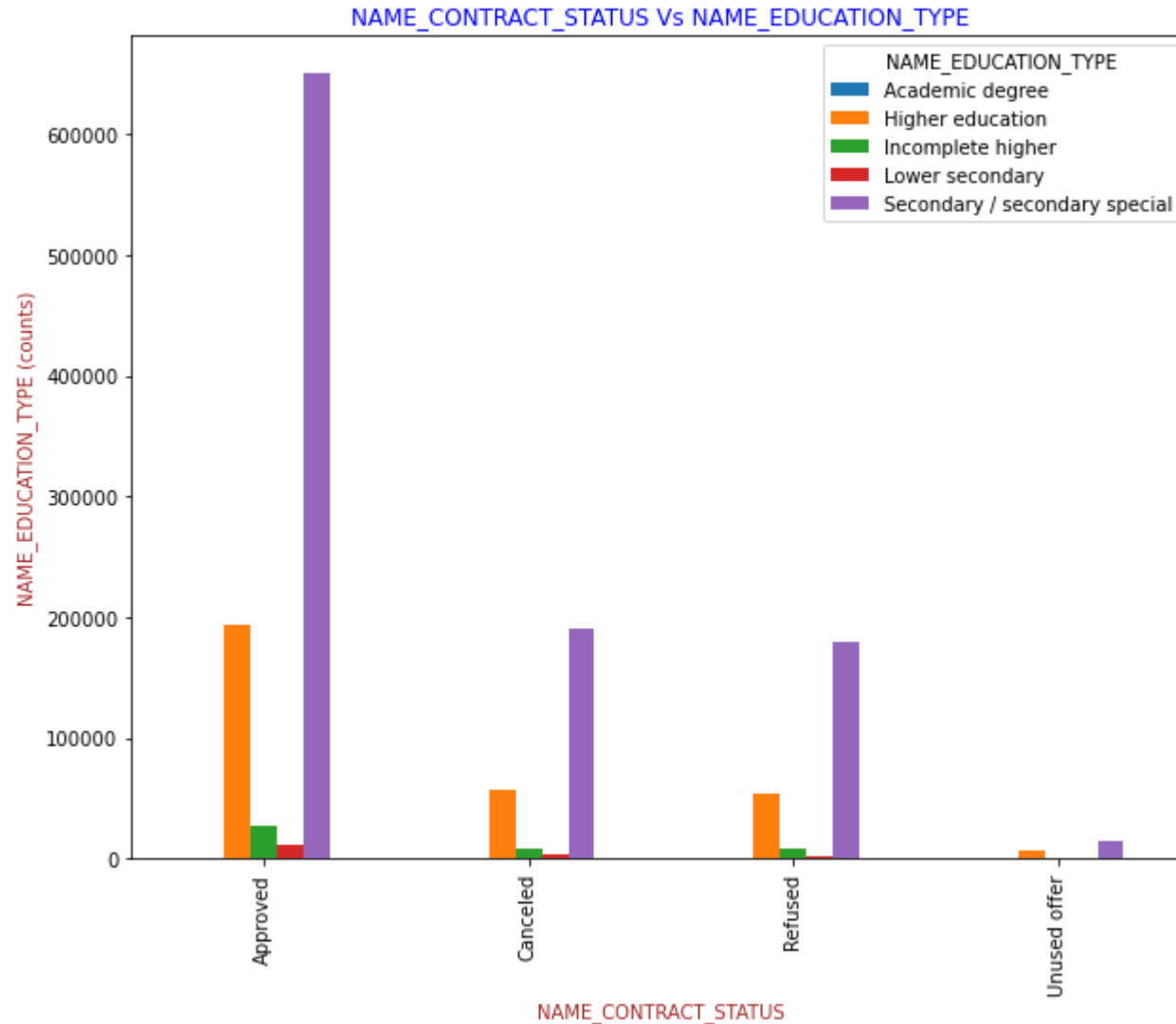
Bivariate analysis on **merged data** – Application data and previous application data.
Occupation Type and Contract Status.

Highest number of approvals for Laborers.

Bivariate analysis on **merged data** – Application data and previous application data.
Education Type and Contract Status.

Highest number of approvals for Secondary/secondary special educated clients.
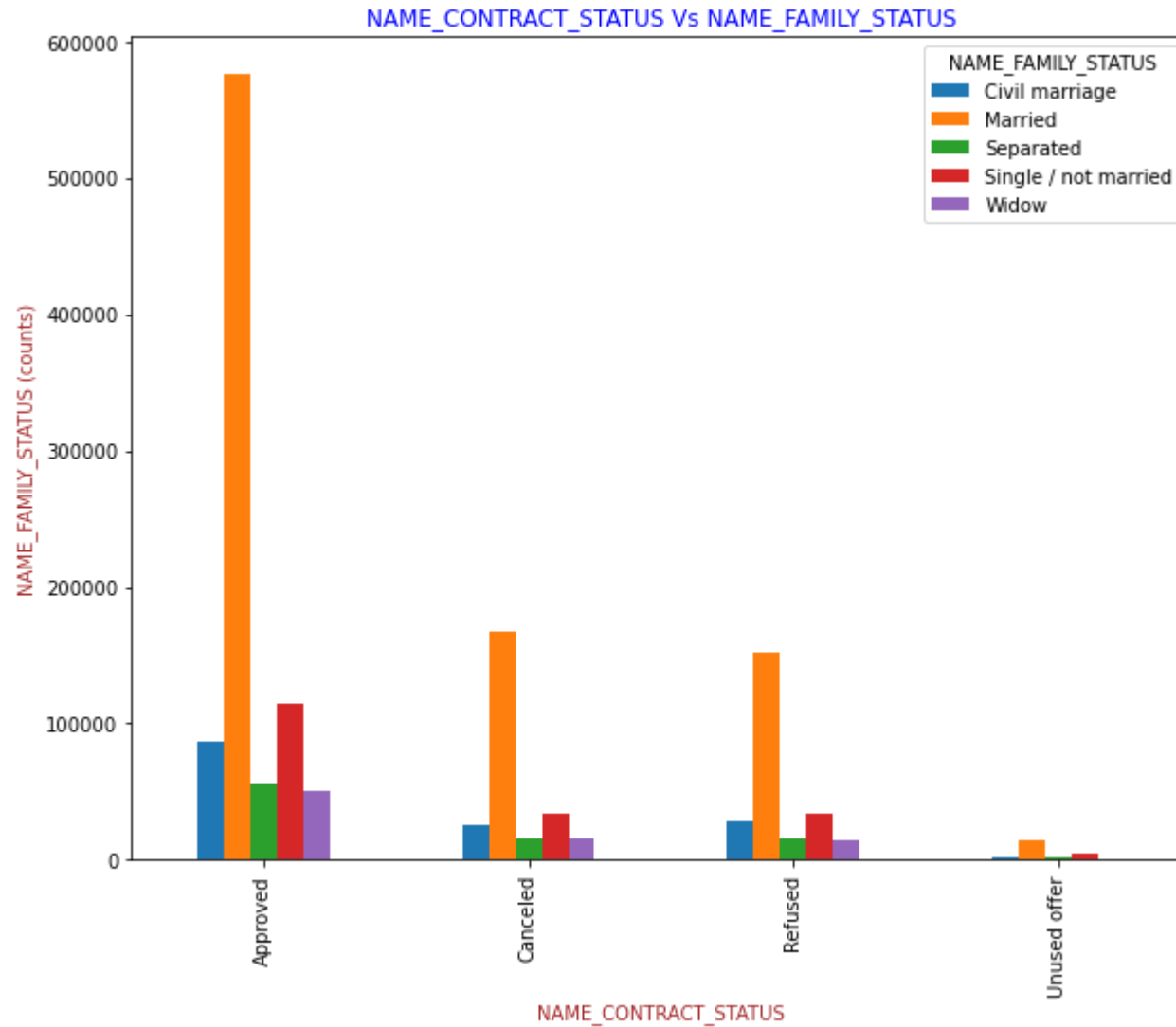
Bivariate analysis on **merged data** – Application data and previous application data.
Family Status and Contract Status.

NAME_CONTRACT_STATUS Vs NAME_FAMILY_STATUS

Highest number of approvals for married clients.

Recommendations:

1. Clients which are in their 20s and 30s, since they are less likely to default.

2. Clients having Occupation such as Security Staff, Cooking Staff, Cleaning Staff, Low-skill Laborers have relatively lower income for all the age group as compared to other occupation types. So we would recommend the with high interest rate, if needed, since they are less reliable.

3. Clients at manager's designation should get the loan credited irrespective of their age.

4. Clients having age more than 60, who are Rental Agents, HRs and Accountants are more credible in terms of loan repayment.

5. Clients having House/apartments are much higher in number and also more likely to default the loans.

6. Self-employed clients and the one who own businesses tend to be in defaulters book.

7. Married clients have highest approval rate as per the previous data. But as seen in application data, the default rate for them is also high.

THANK YOU!