CREDIT EDA ASSIGNMENT

By Soumyashree Behera
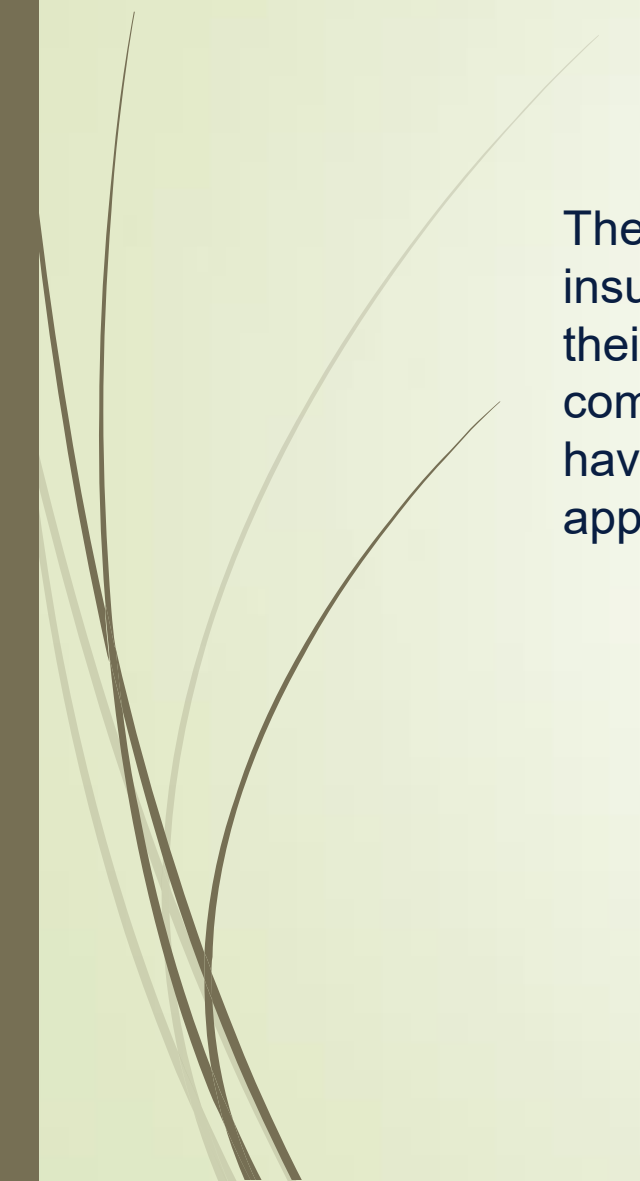
Content

- Problem Statement
- Approach Used
- Steps Followed
- Data Analysis Reports
- Conclusions

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 to 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 have to use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

Approach Used

- Analyzing and understand the application_data.csv dataset which contains all the information of the client at the time of application.

- Perform cleaning of the 1st dataset by applying various techniques.

- Analyzing and understand the previous_application.csv dataset which contains information about the client's previous loan data.

- Perform cleaning of the 2nd dataset by applying various techniques.

- Merge both of the above data set and understand the compete dataset.

Steps Followed

- Understanding the variables.
- Import the data sets.
- Verify the structure of the data set.
- Validate the missing value.
- Dropping irrelevant columns.
- Validation of outlier.
- Imputing the values (if needed).
- Performing Univariant Analysis.
- Performing Segmented Univariant Analysis.
- Performing Bivariant Analysis.

Data Analysis Reports

Data Imbalance based on Gender

- In the chart we see that percentage of females is more.

- So more females have applied for the loan.

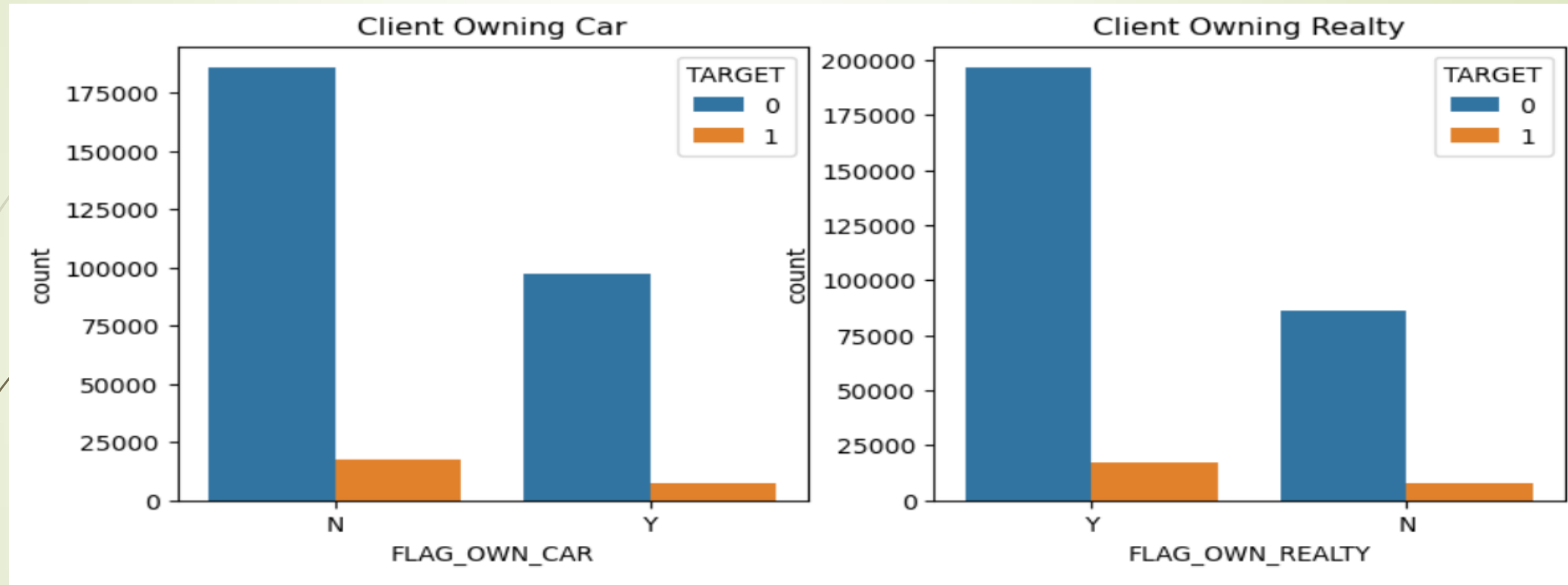Loan applied based on Gender
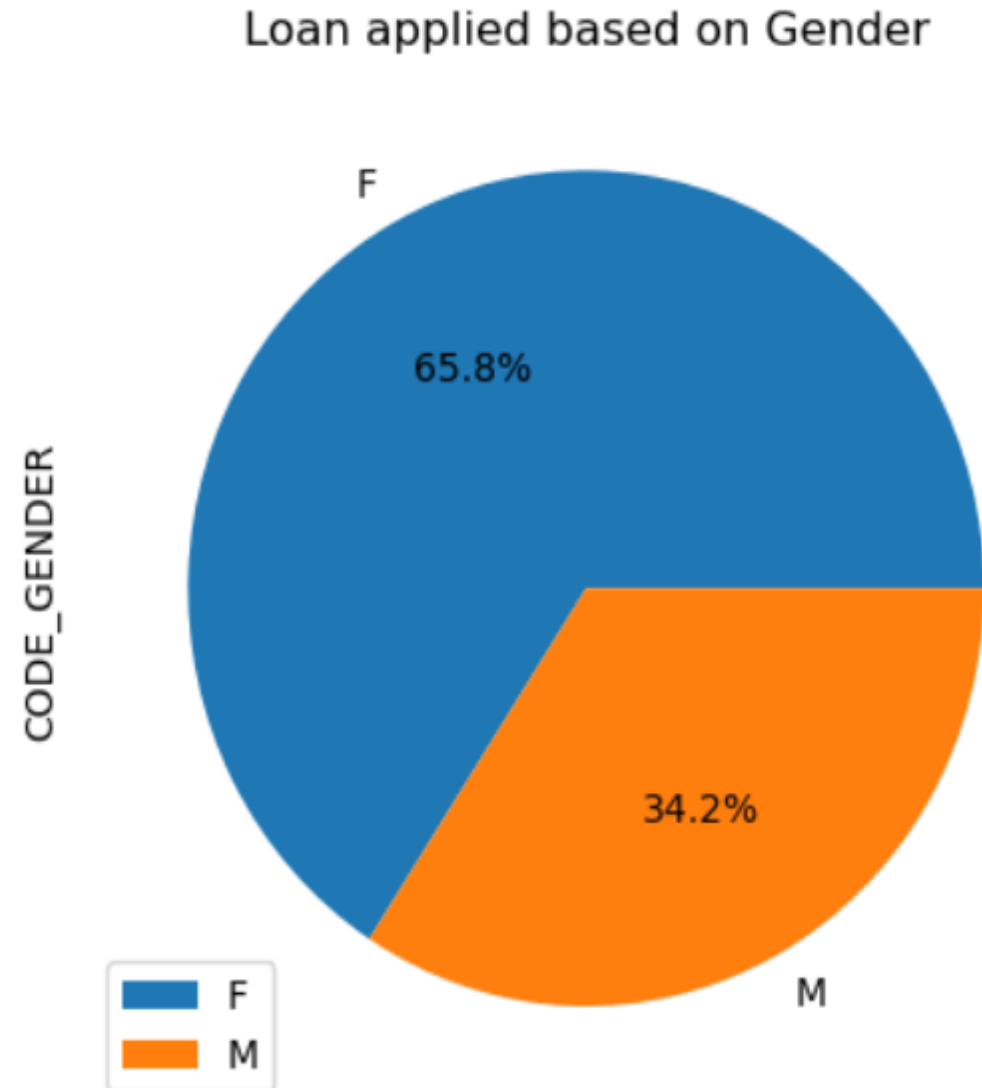
Data Imbalance based on Target

- In the chart we see that percentage of clients paying loan properly is more.

- Hence more clients are non-defaulters.

- Clients with pay difficulties are less as compared to clients will no payment difficulties.
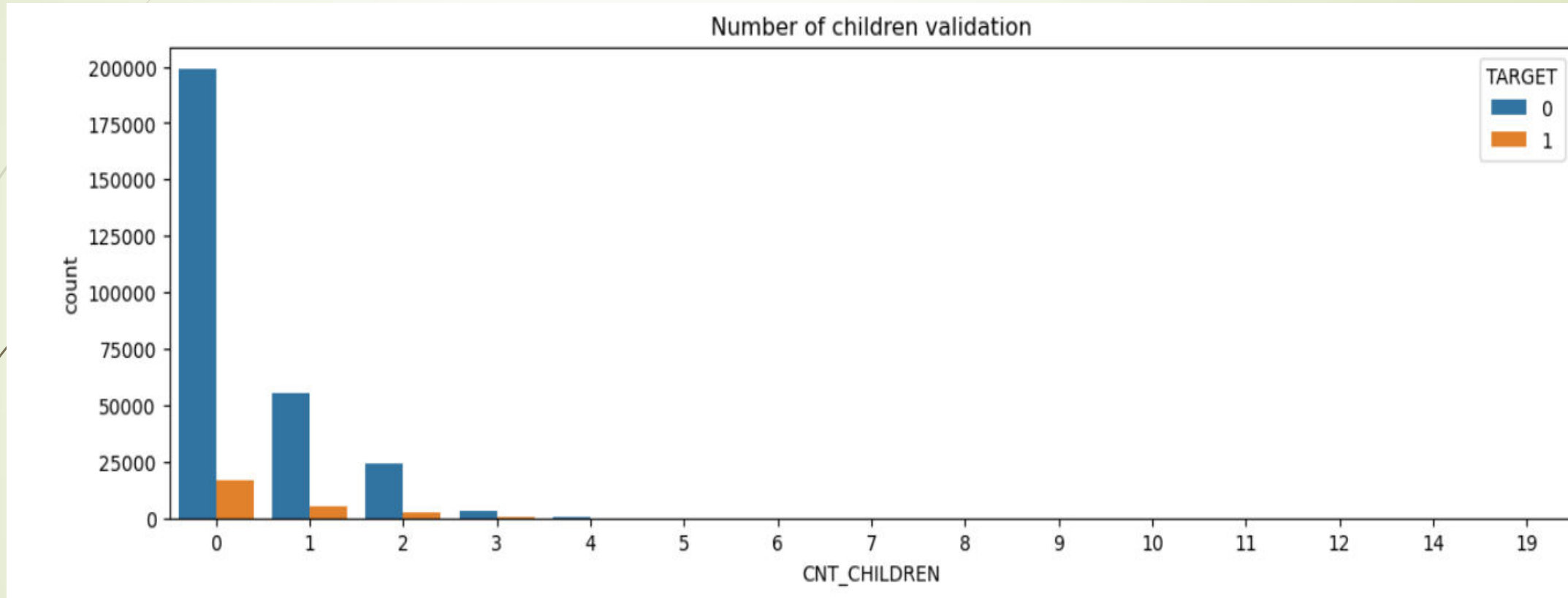

Clients based on Target
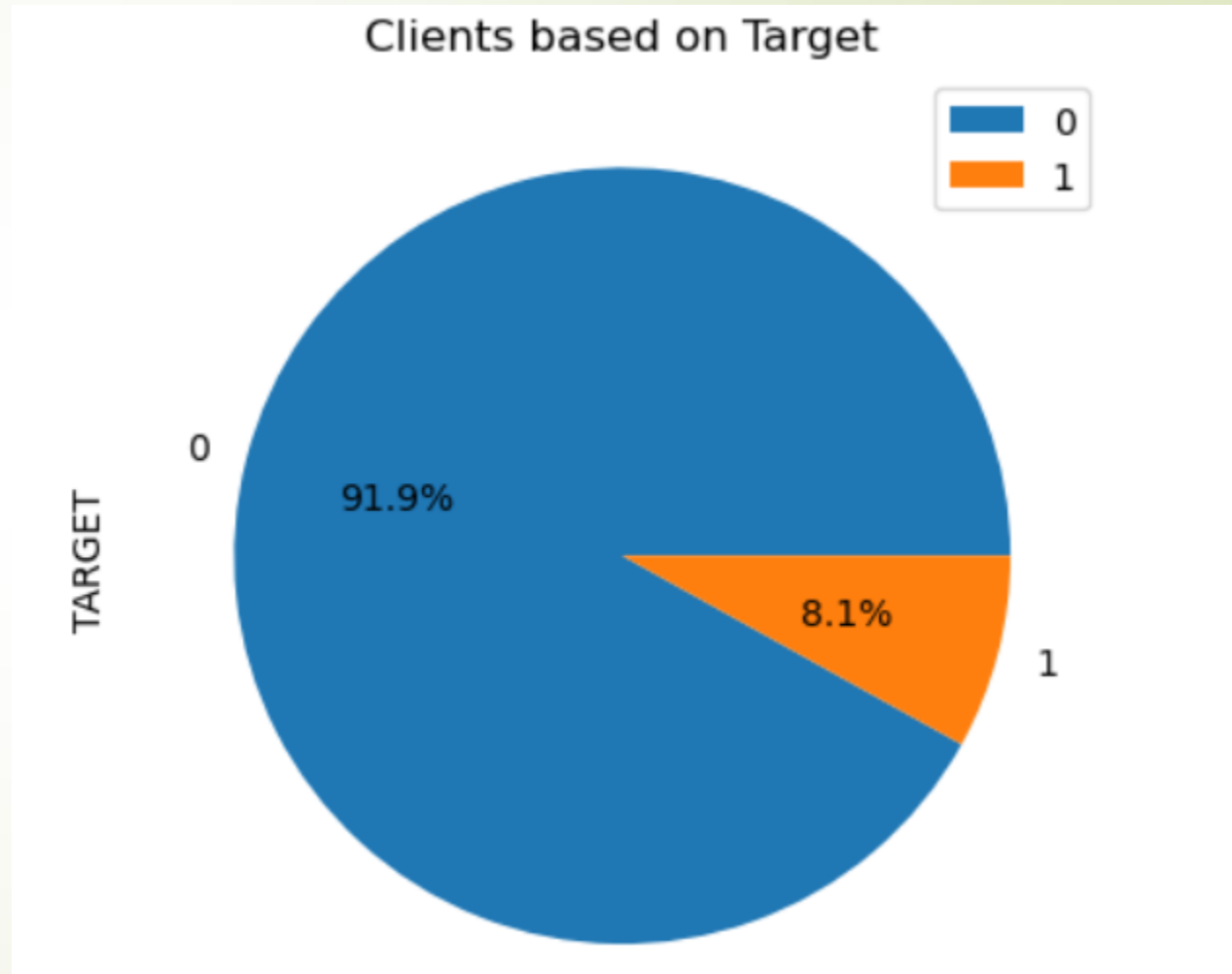
Clients owning car vs Clients owning house

- Considering the above chart Clients who don't have car are paying the loan properly.
- Clients who own the House are paying the loan properly.
- Hence Clients who own house but not car have no payment difficulties.

Analysis based on CNT_CHILDREN

Number of children validation

- Considering the above chart, zero children are more for the one's who paid loan properly.
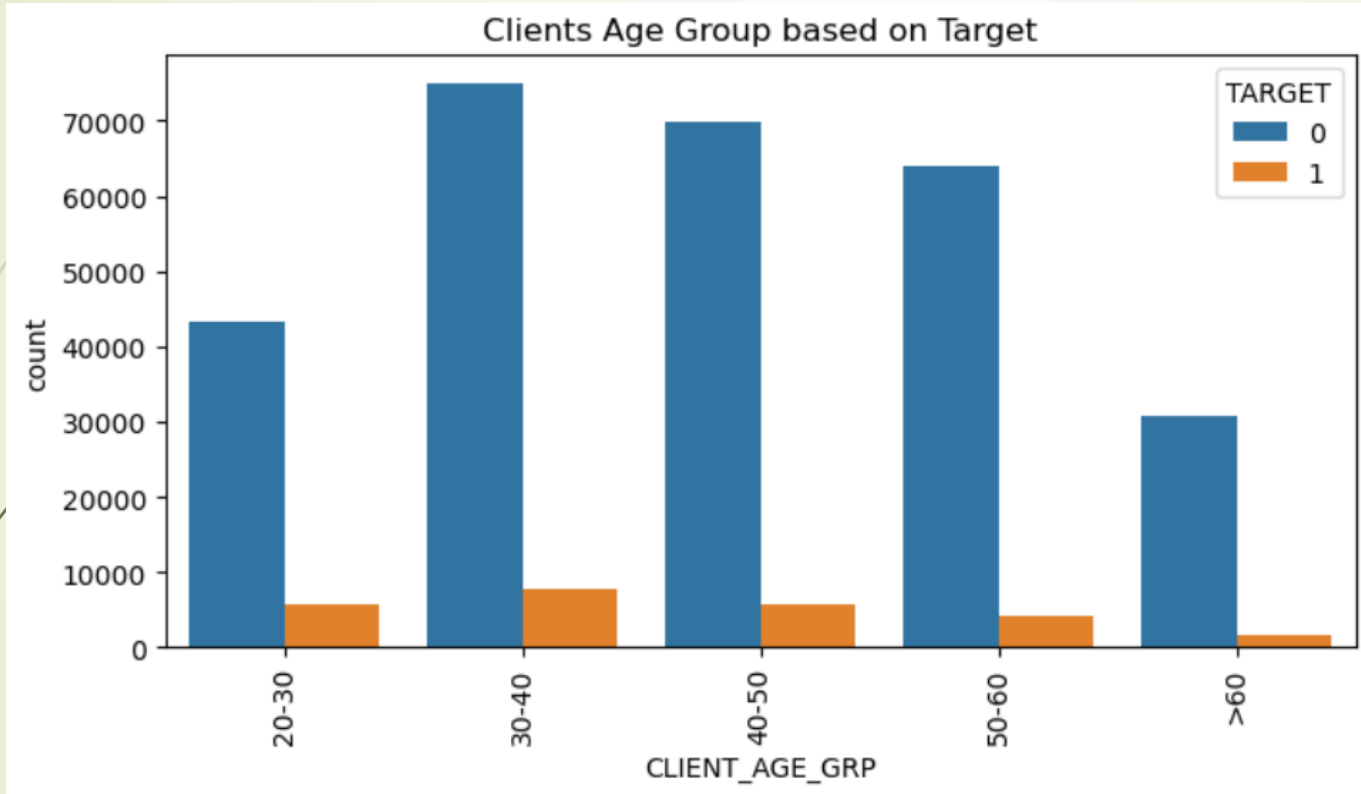- Hence, clients having no children are paying the loan properly.

Clients Age Based on Gender

| CODE_GENDER	F	M
TARGET		
0	45.483001	41.689978
1	41.646366	39.625434

Clients Age Based on Gender
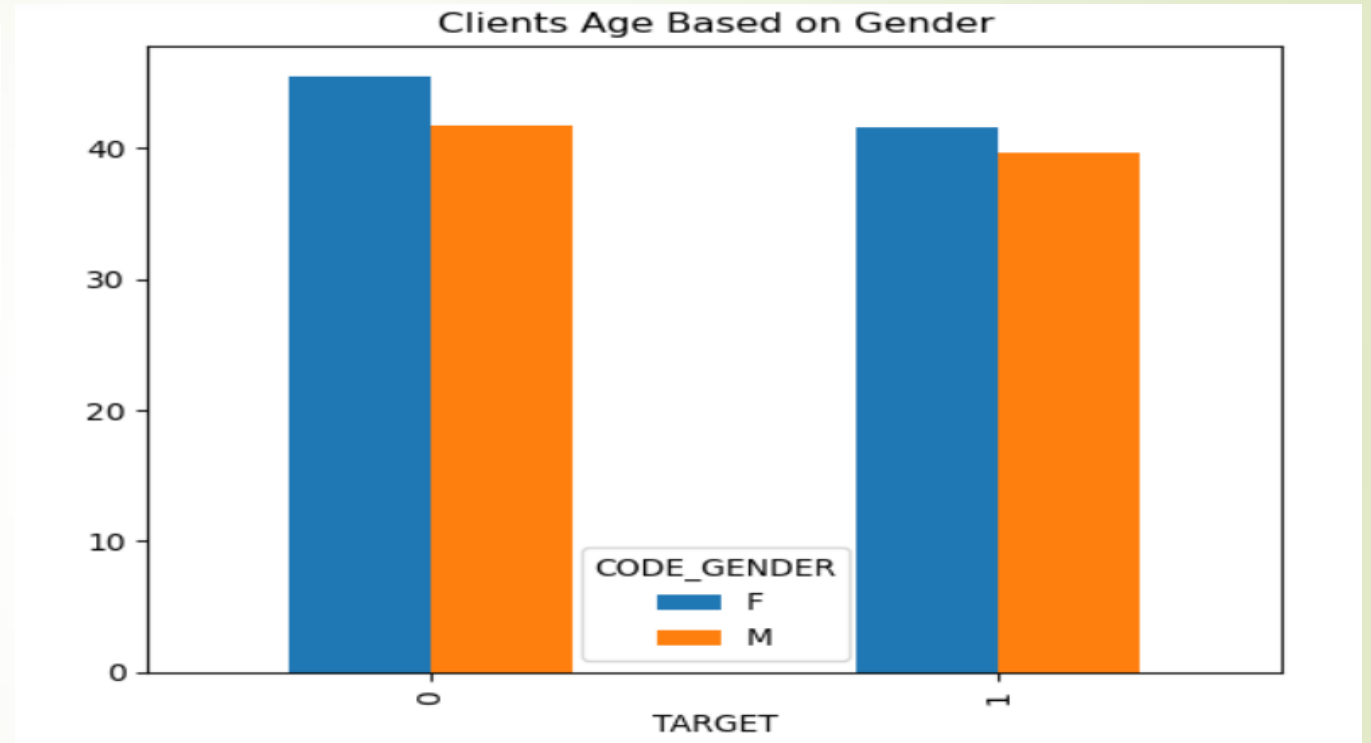
- Female population are more who are paying the loan properly.

Clients Age Group based on Target

- Based on the above chart, we see that 30-40 range have more clients who pays the loan properly.

- Hence, majority of clients with age group of 30-40 are paying the loan properly.

Type of Income based on Target
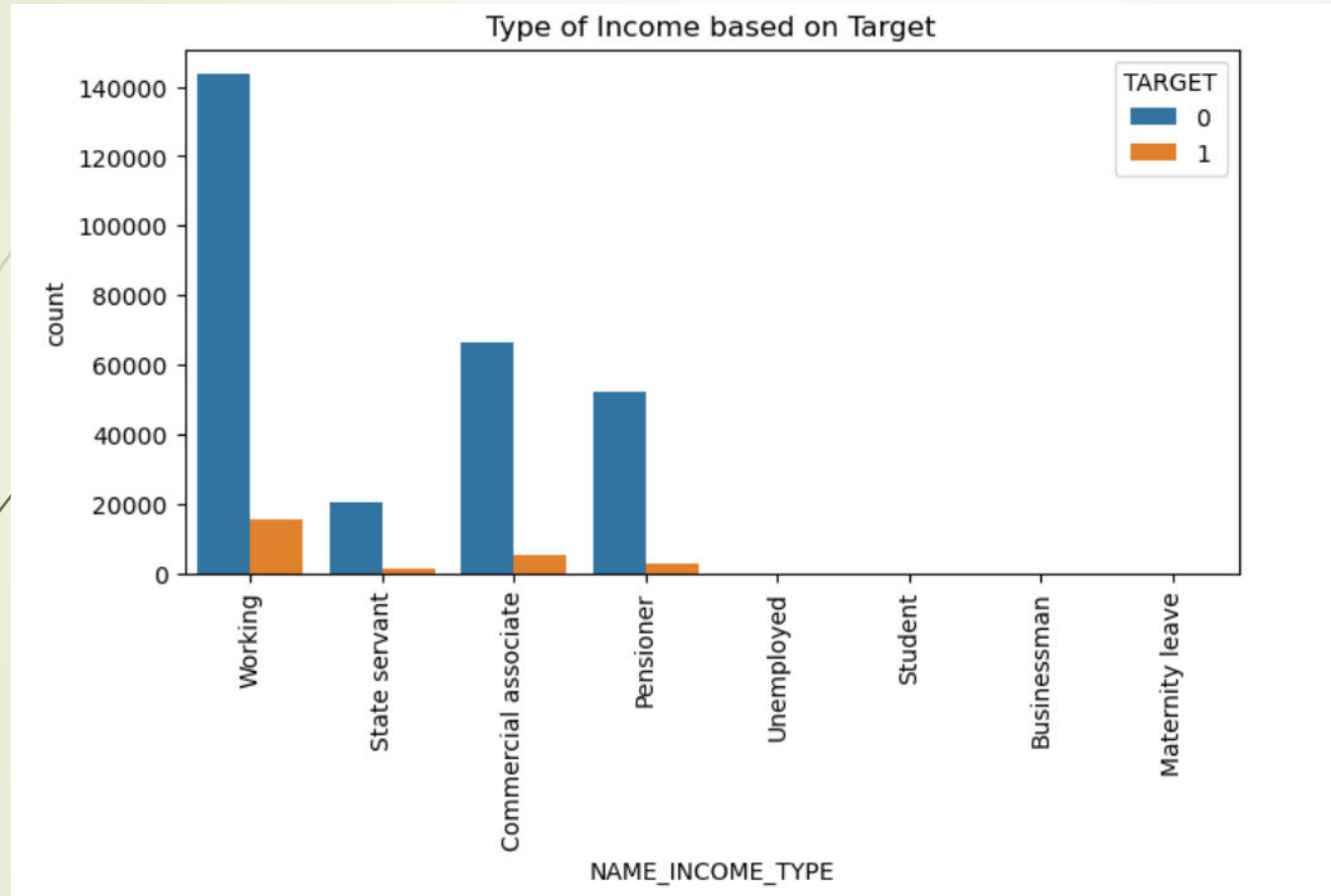
- Based on the above chart, we see that working clients are more.
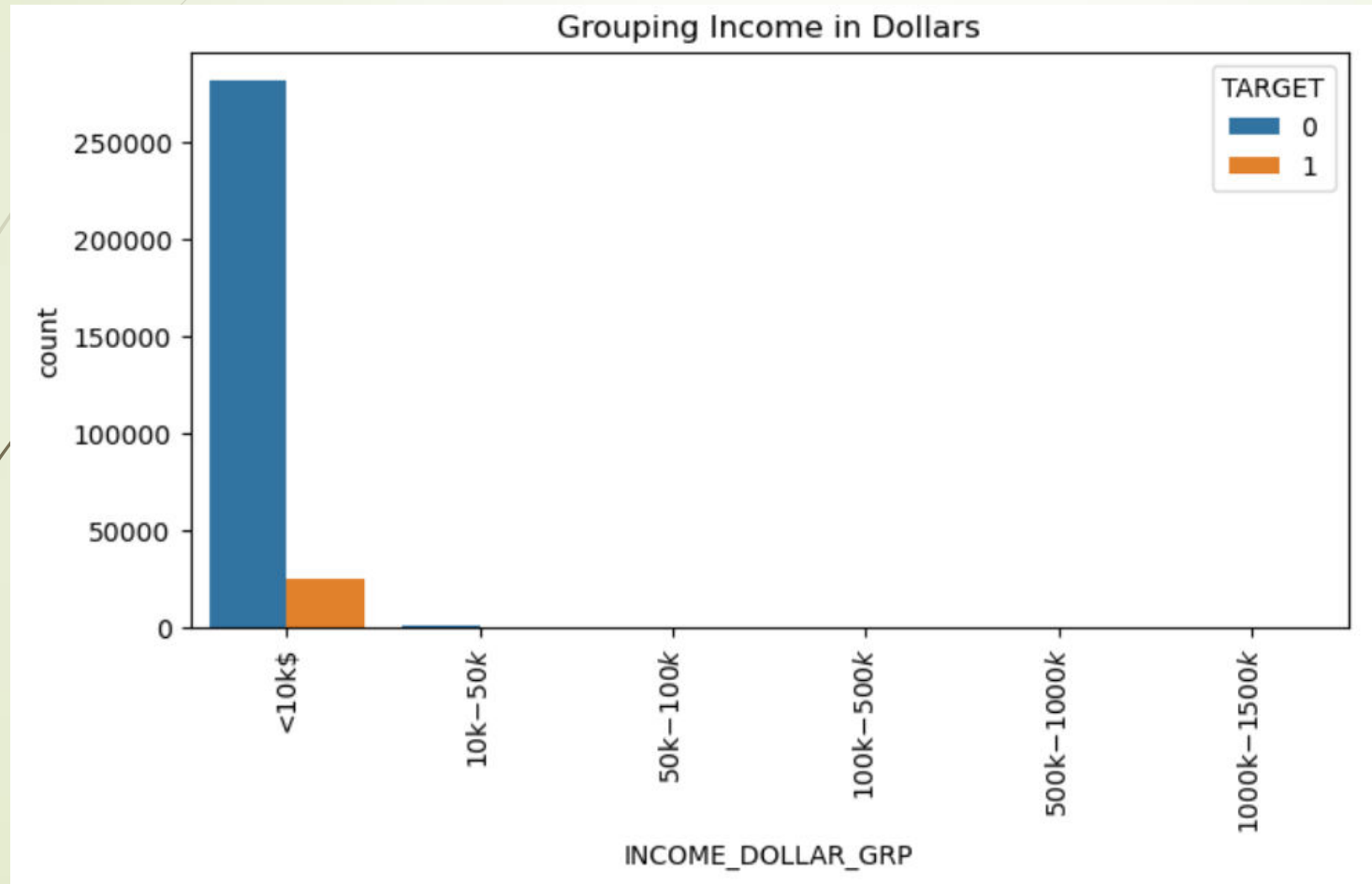- Hence, working peoples are paying the loans properly.

Clients Education
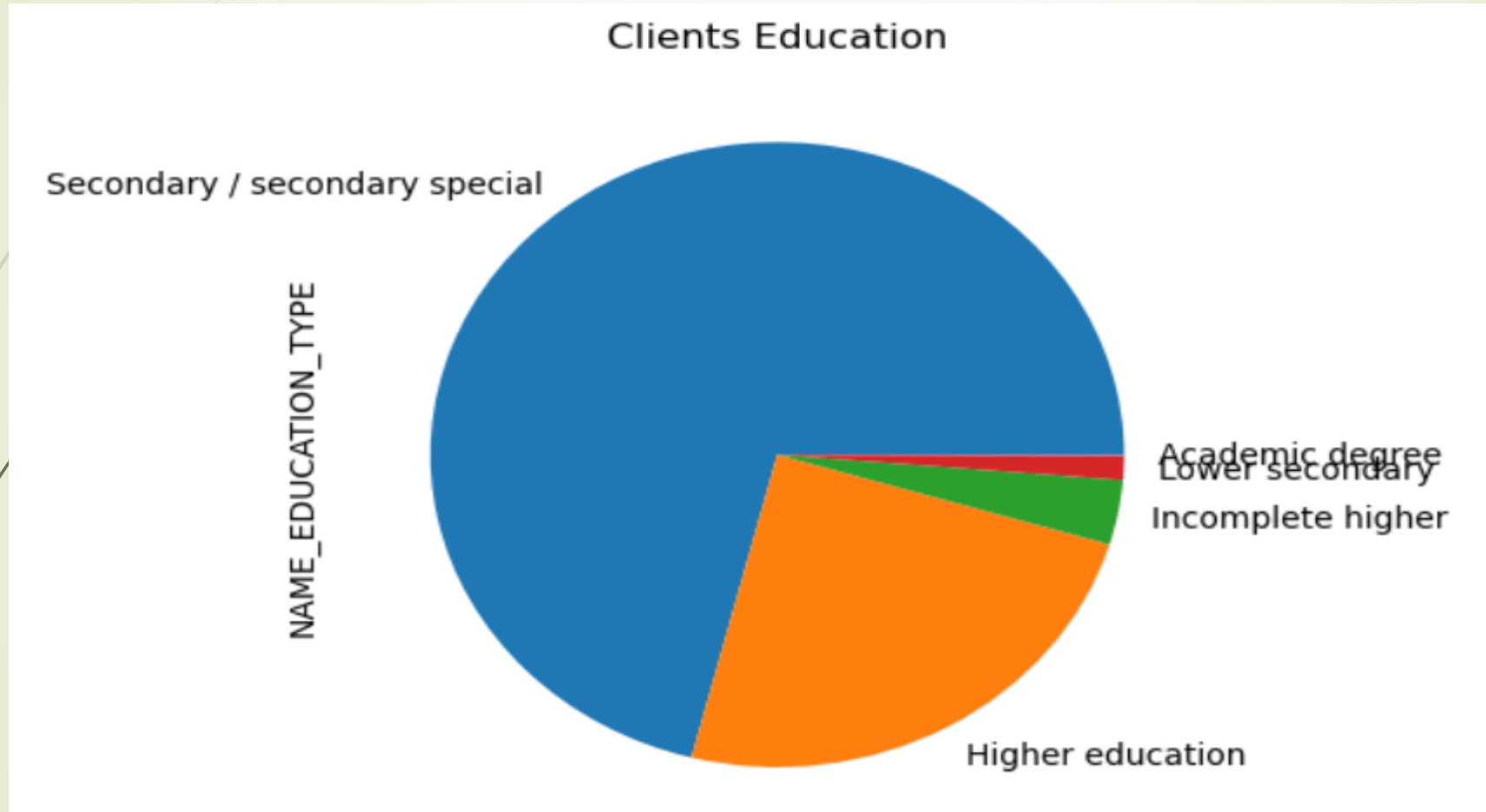
Based on the above chart we see that clients who have education type as 'Secondary/Secondary Special' have taken more loans.

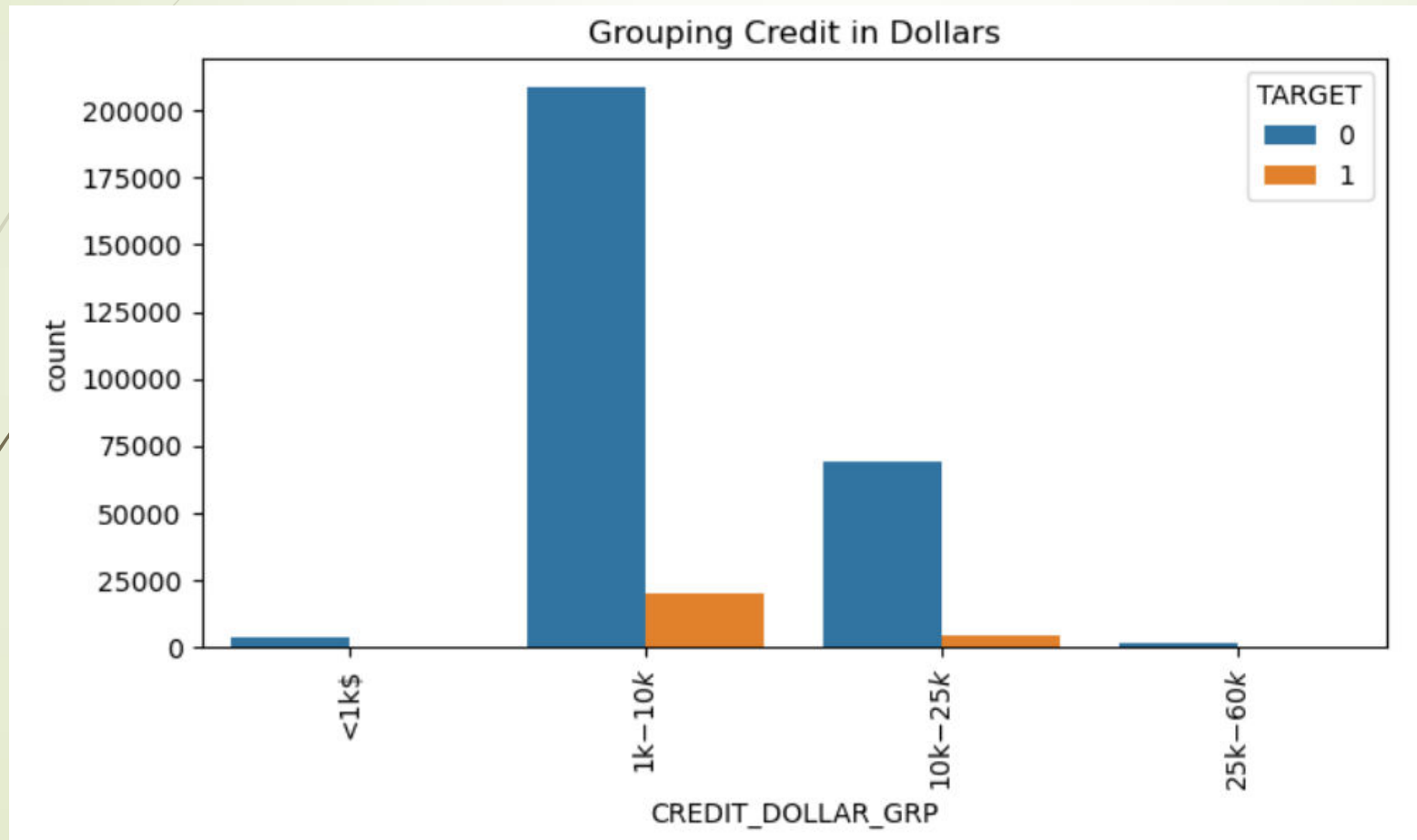# Clients Income Range

NOTE: Amount is converted to dollar (1dollar=81 rupees)

Grouping Income in Dollars

- Based on the above chart we see that clients having Income range less than 10000 dollar are the one's who were paying of loans properly.

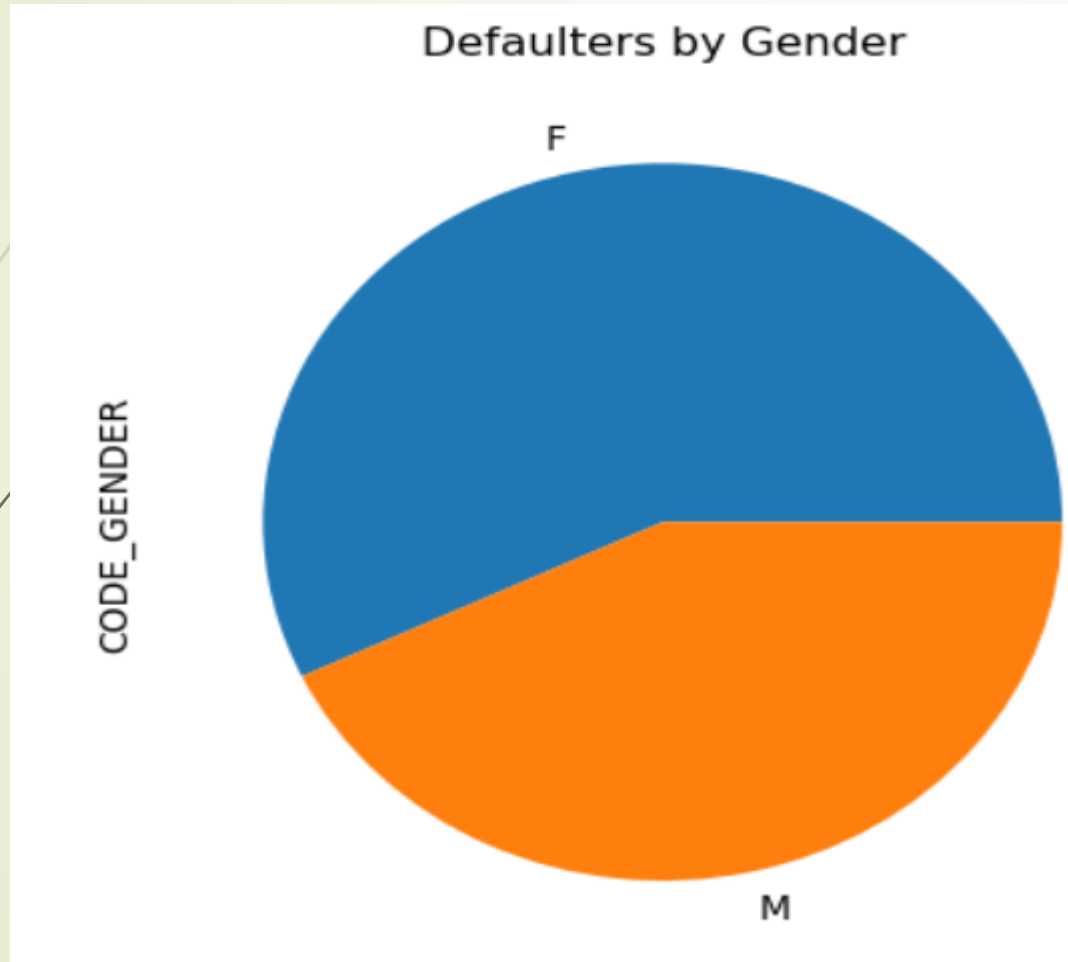# Grouping Credit in Dollars

NOTE: Amount is converted to dollar (1dollar=81 rupees)

- Based on the above chart we see that many clients have taken credits between 1000 to 10000 dollars are the one's who were paying off the loan properly.
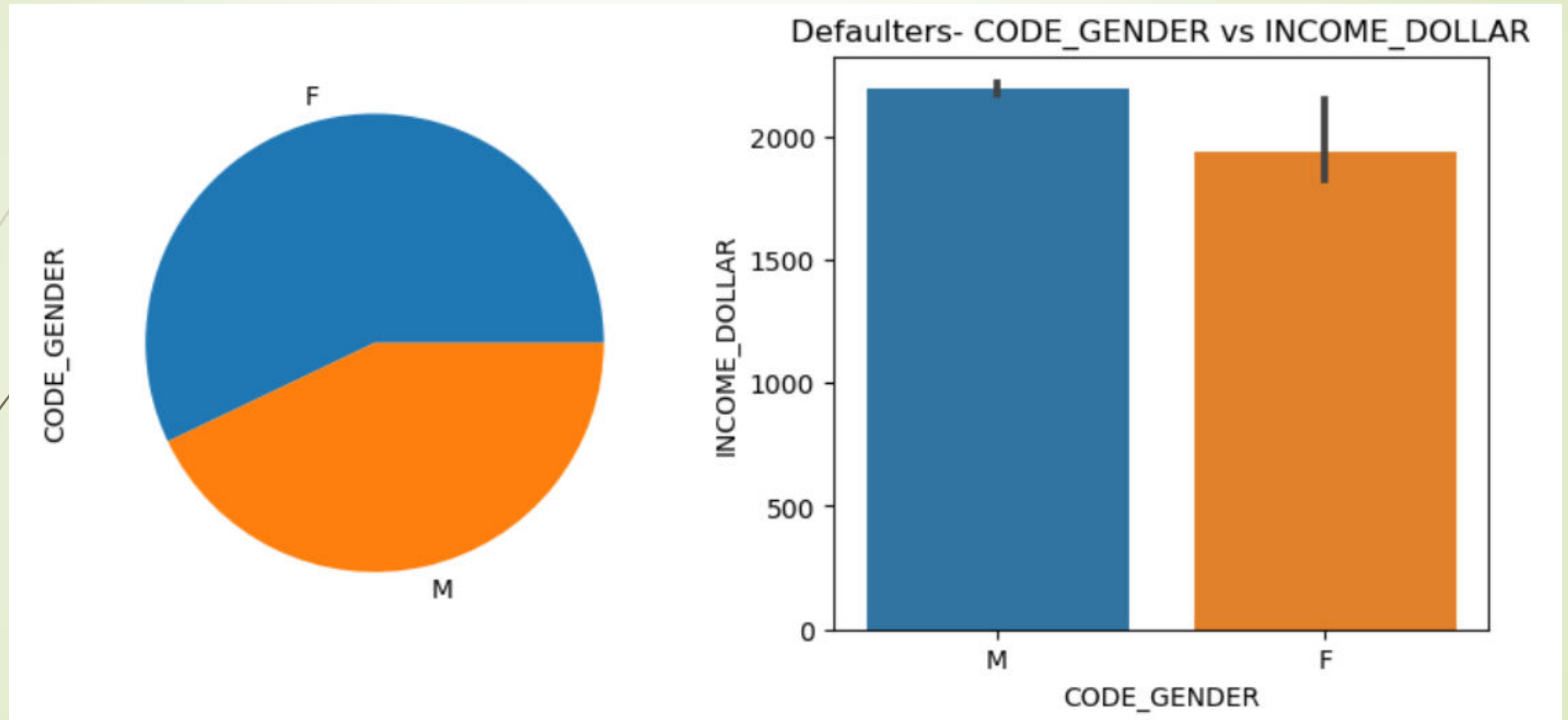
Defaulters by Gender

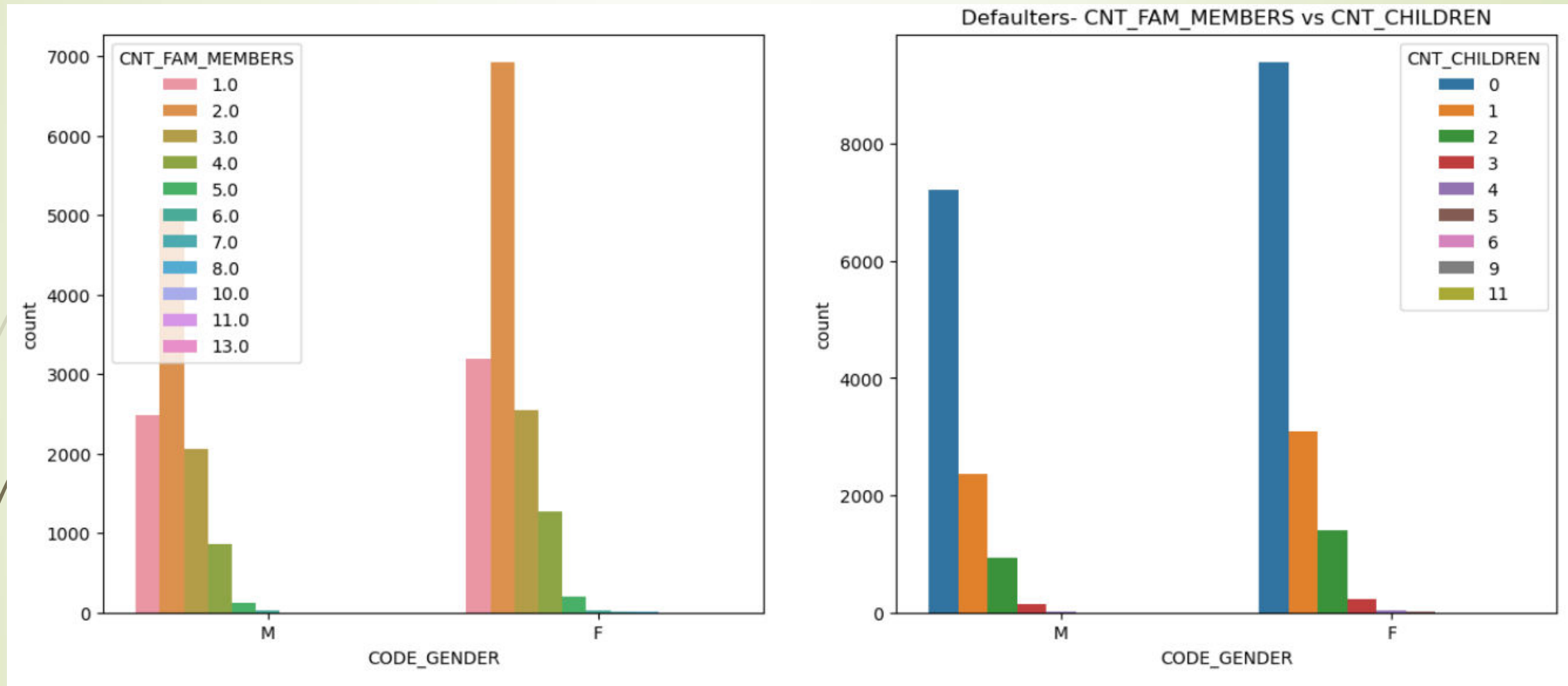- Based on the above chart we see that more females compared to male are having difficulties in paying the loan.

Defaulters- CODE_GENDER vs INCOME_DOLLAR

- Based on the above chart we see that males who income more are the defaulters and females who paid less have difficulties in paying off the loans.

Defaulters- CNT_FAM_MEMBERS & CNT_CHILDREN columns

Defaulters- CNT_FAM_MEMBERS vs CNT_CHILDREN

- Based on the above chart we see that clients having 2 family members without children are the defaulters.

Defaulters: Income Amount vs Credit
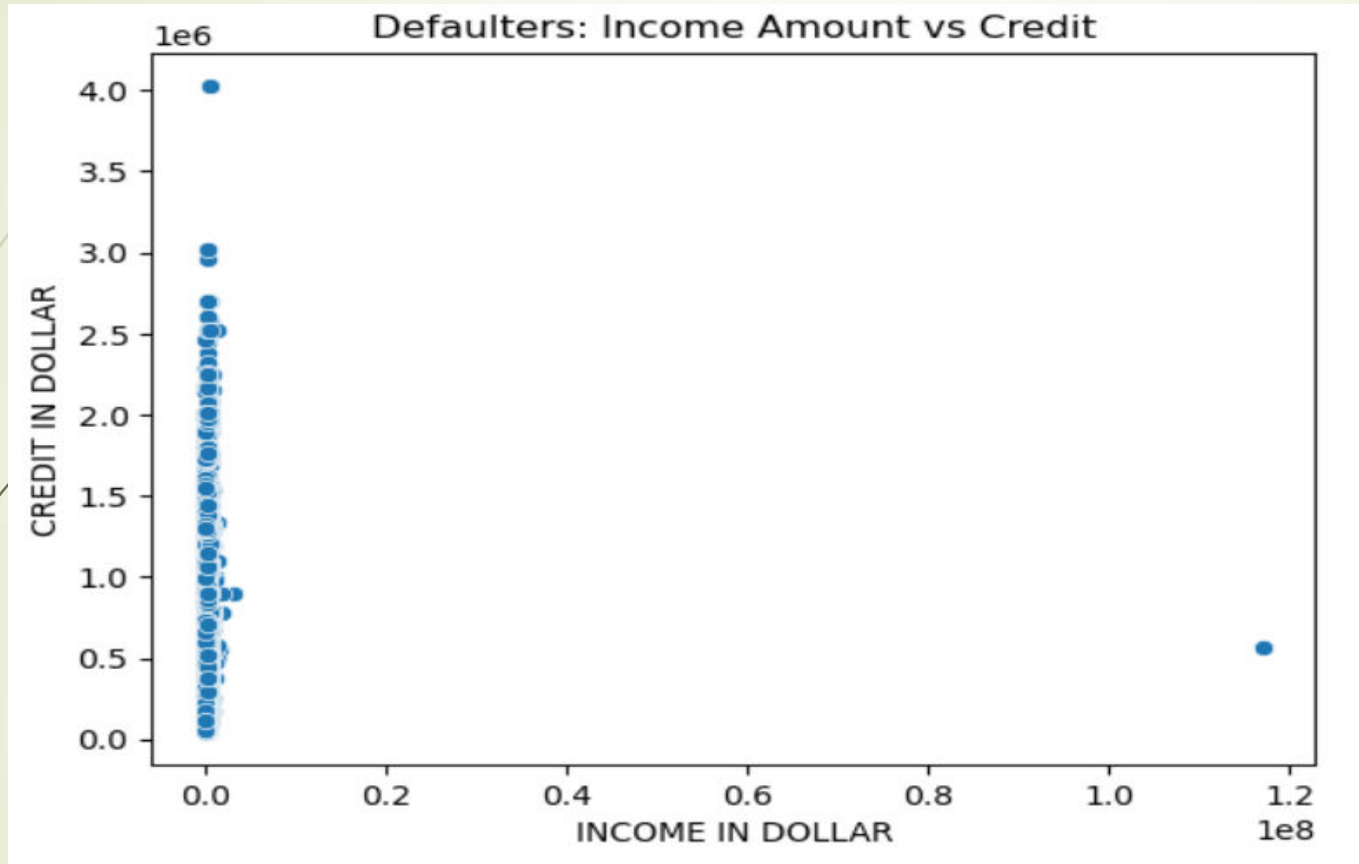
- Based on the above chart we see that clients with low income range peoples are more defaulters.

Defaulters: Family Status and Gender based on Credit Dollars

NAME_FAMILY_STATUS	Civil marriage	Married	Separated	Single	Widow
CODE_GENDER					
F	6271.051398	7400.636768	6605.296967	5963.264563	6555.370785
M	6206.232276	7334.428820	6405.828557	5861.552620	5749.069541

- Based on the above chart we see that married male and married female were having difficulties in paying the loan.

Merged Dataset- Clients respective to Contract Status
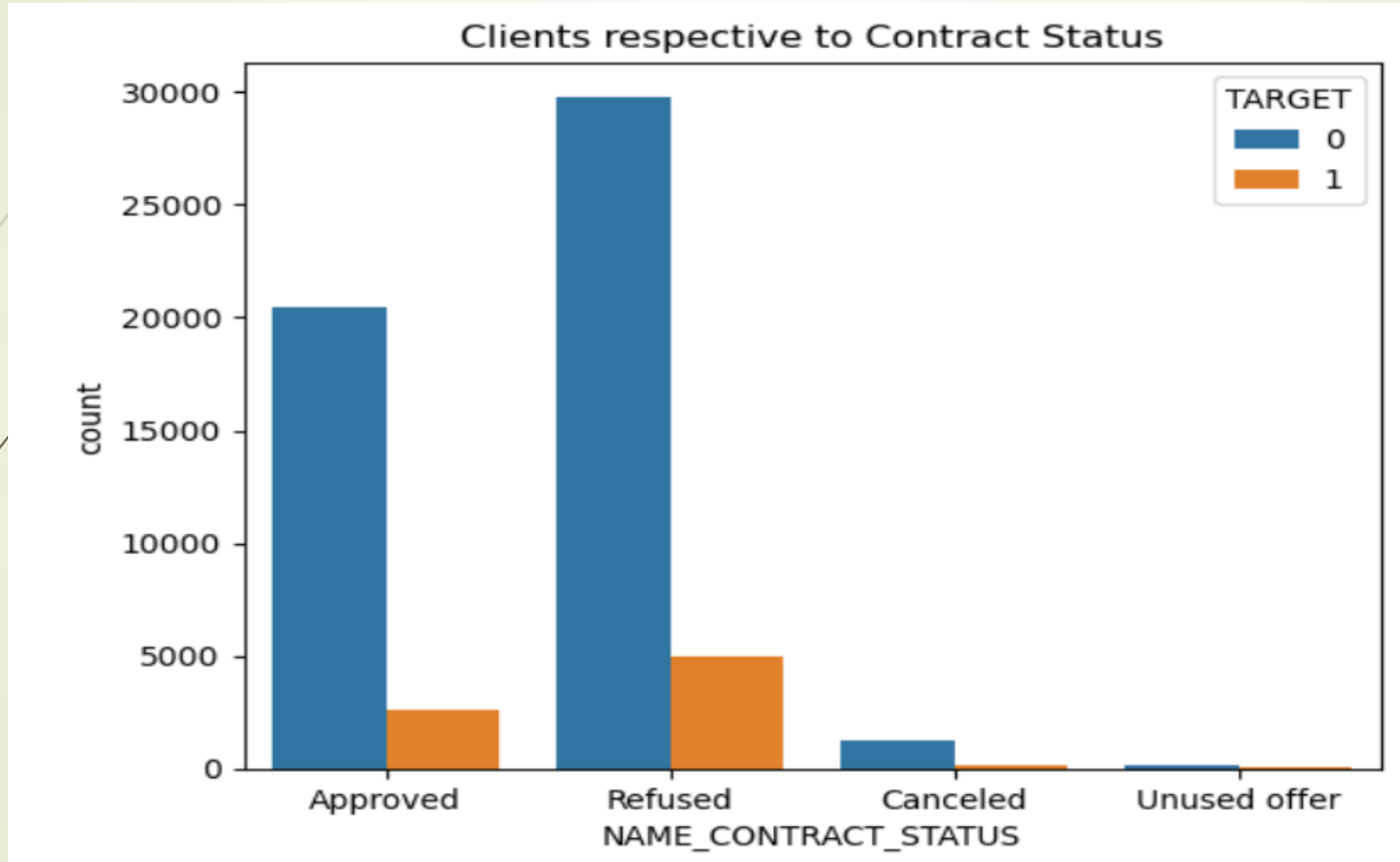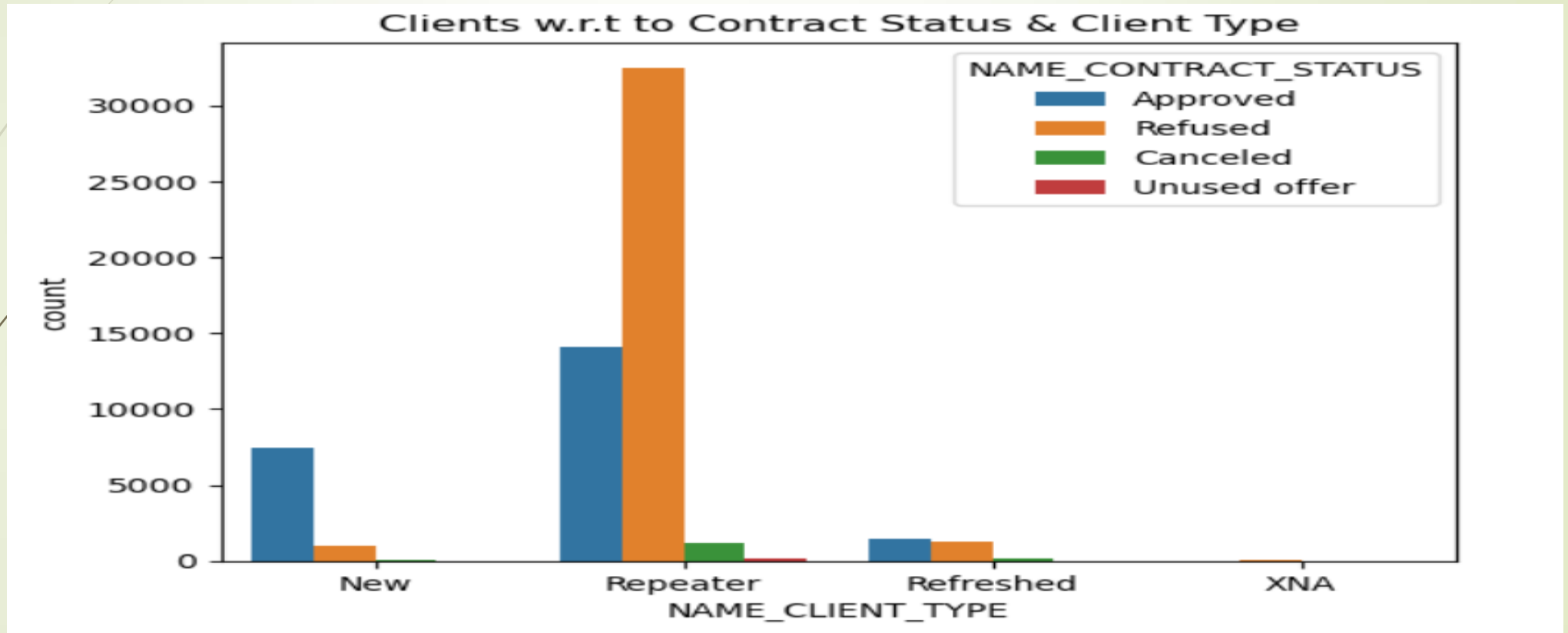

Clients respective to Contract Status

- Based on the above chart we see that clients whose previous application was refused are the most defaulters.

Merged Dataset- Clients w.r.t to Contract Status & Client Type

■ Based on the above chart we see that repeaters client are more whose loan is refused.

Conclusions

- More Females have applied for loan.
- Clients with pay difficulties are less as compared to clients will no payment difficulties.
- Clients who own house but not car have no payment difficulties.
- Clients having no children are paying the loan properly.
- Female population are more who are paying the loan properly.
- Majority of clients with age group of 30-40 are paying the loan properly.
- Working peoples are paying the loans properly.
- Clients who have education type as 'Secondary/Secondary Special' have taken more loans.
- Clients having Income range less than 10000 dollar are the one's who were paying of loans properly.
- Many clients have taken credits between 1000 to 10000 dollars are the one's who were paying off the loan properly.
- More Females compared to men are having difficulties in paying the loan.
- Males who income more are the defaulters and females who paid less have difficulties in paying off the loans.
- Clients having 2 family members without children are the defaulters.
- Clients with low Income range peoples are more defaulters.
- Married Male and Female were having difficulties in paying the loan.
- Clients whose previous application was refused are the most defaulters.
- Repeaters client are more whose loan is Refused.