

Welcome Everyone



Credit EDA Assignment



By - Ruchi Dham

Problem Statement

- ▶ The Problem is related to Finance and Banking Services which provide loans.
- ▶ The dataset has multiple columns related to the information about the customers.
- ▶ This dataset to be analyzed and factors to be identified which helps in finding the defaulters and good customers.
- ▶ Bank don't want to give loan to defaulters.
- ▶ Bank want to give loan to good customers.

Data Understanding

- ▶ *'application_data.csv'* contains all the information of the client at the time of application. The data is about whether a **client has payment difficulties**.
- ▶ *'previous_application.csv'* contains information about the client's previous loan data. It contains the data on whether the previous application had been **Approved, Cancelled, Refused or Unused offer**.
- ▶ *'columns_description.csv'* is data dictionary which describes the meaning of the variables.

Objective

- ▶ Perform EDA process on the datasets.
- ▶ Find some useful insights, patterns and variable which clear indicates.
- ▶ User is likely to default the loan or having payment difficulties.
- ▶ User will pay their installments on time.
- ▶ Insights from these datasets to help bank for identifying the defaulters to mitigate the business loss and good customers to have good business.

Approach

► Data Cleaning

1. Deleted those columns which have more than 50% of null data.
2. Deleted the columns which are having no impact on our data and of no use in analyzation.

► Data Correction and modification

1. Changed the days value to positive and converted them in years.
2. Corrected the data types of some columns.

Methods

► Outliers

Find the outliers in numerical columns and suggested some methods to handle it as it was not suggested us to handle the outliers.

► Binning

Perform the binning in some columns arrange them in specific manner.

► Merging

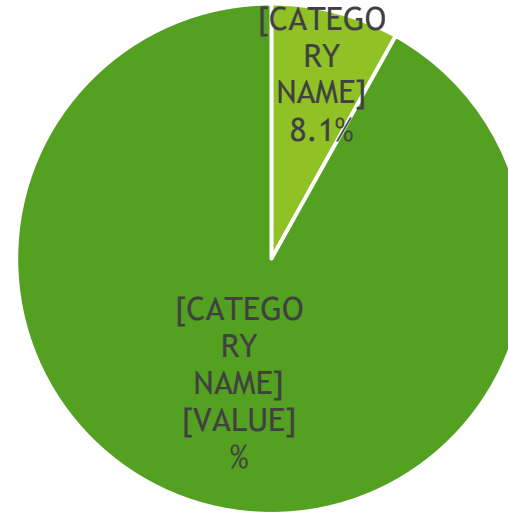
Merged the two data frames on SK_ID_CURR.

► Analyse

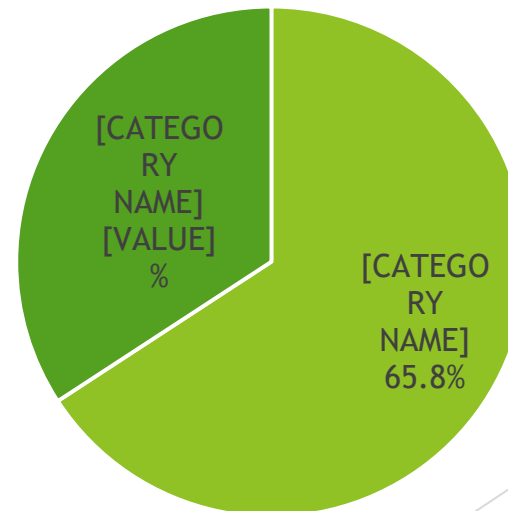
Perform the analysis on these data.

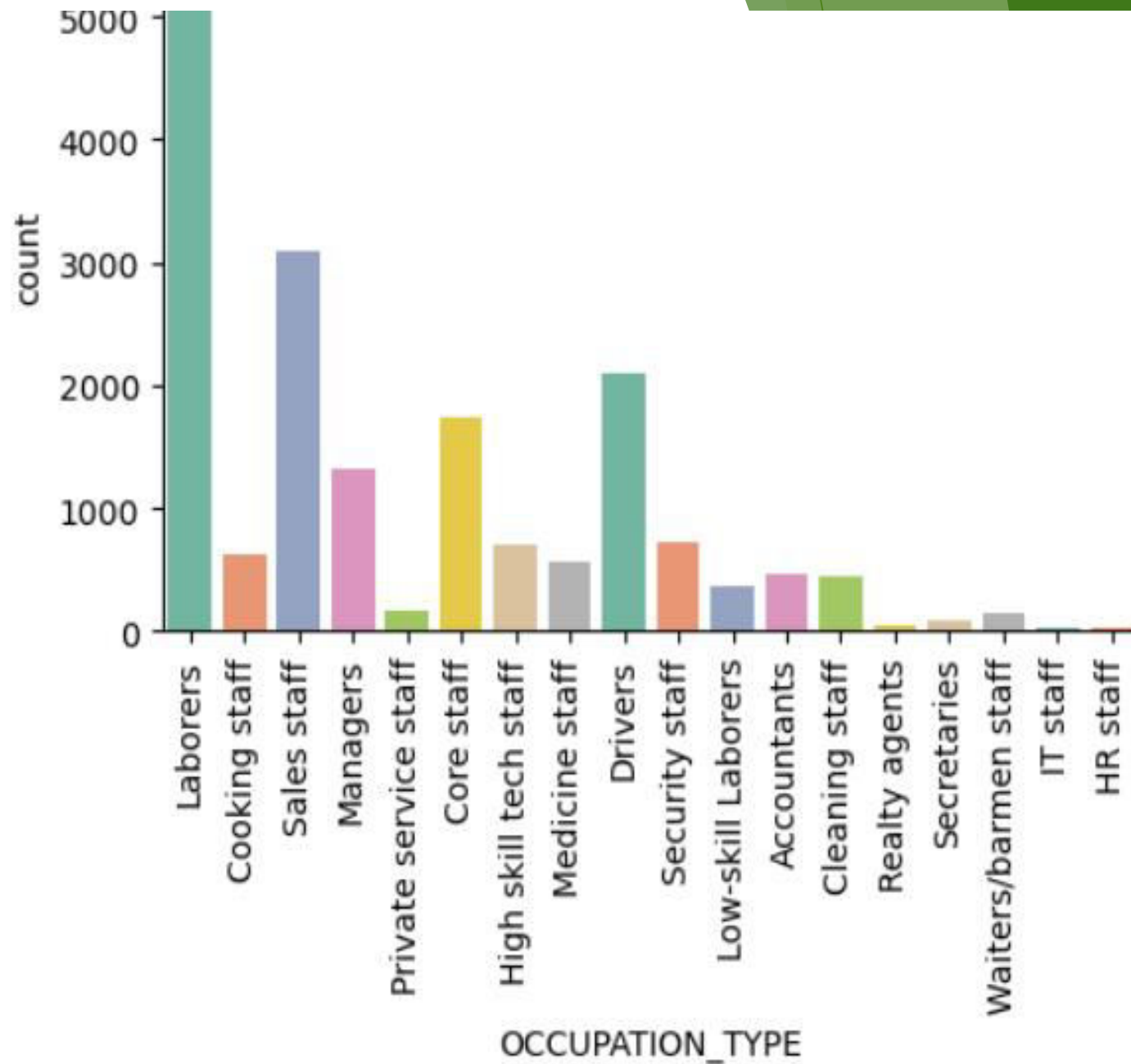
Insights

- ▶ Only 8.1% customers are defaulters the rest are non-defaulters.



- ▶ There are 65.8% females who are applying for loan.

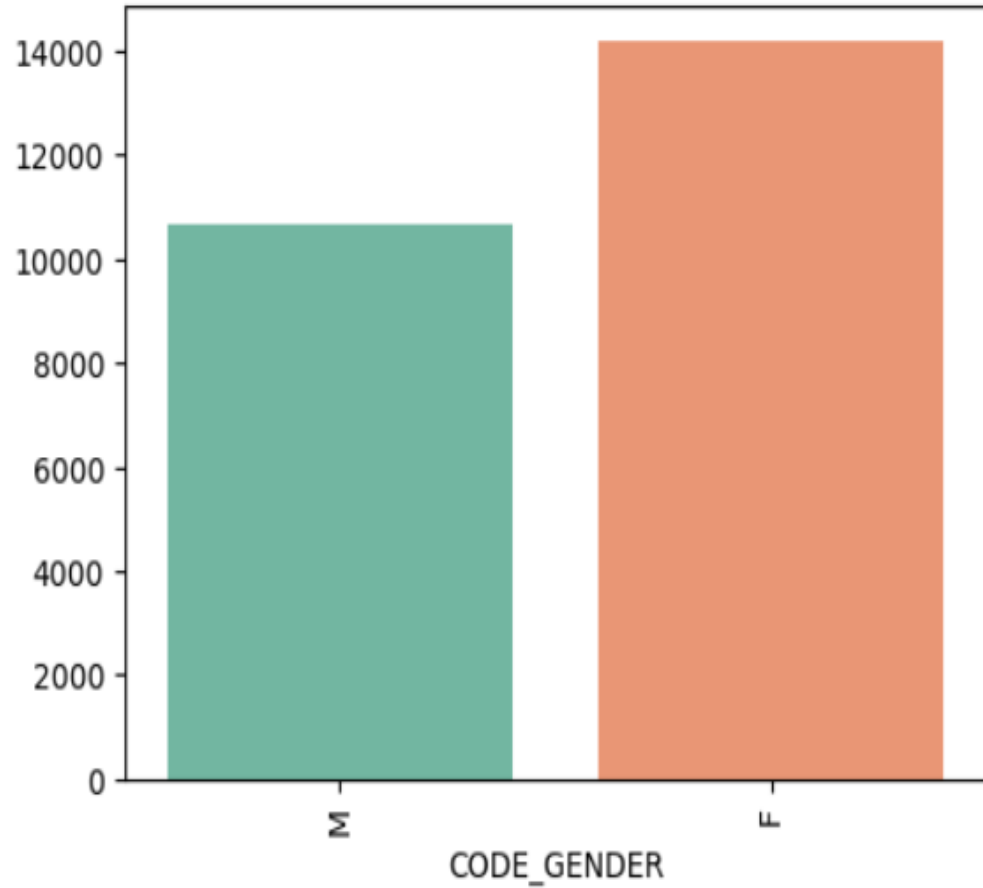




Occupation Type

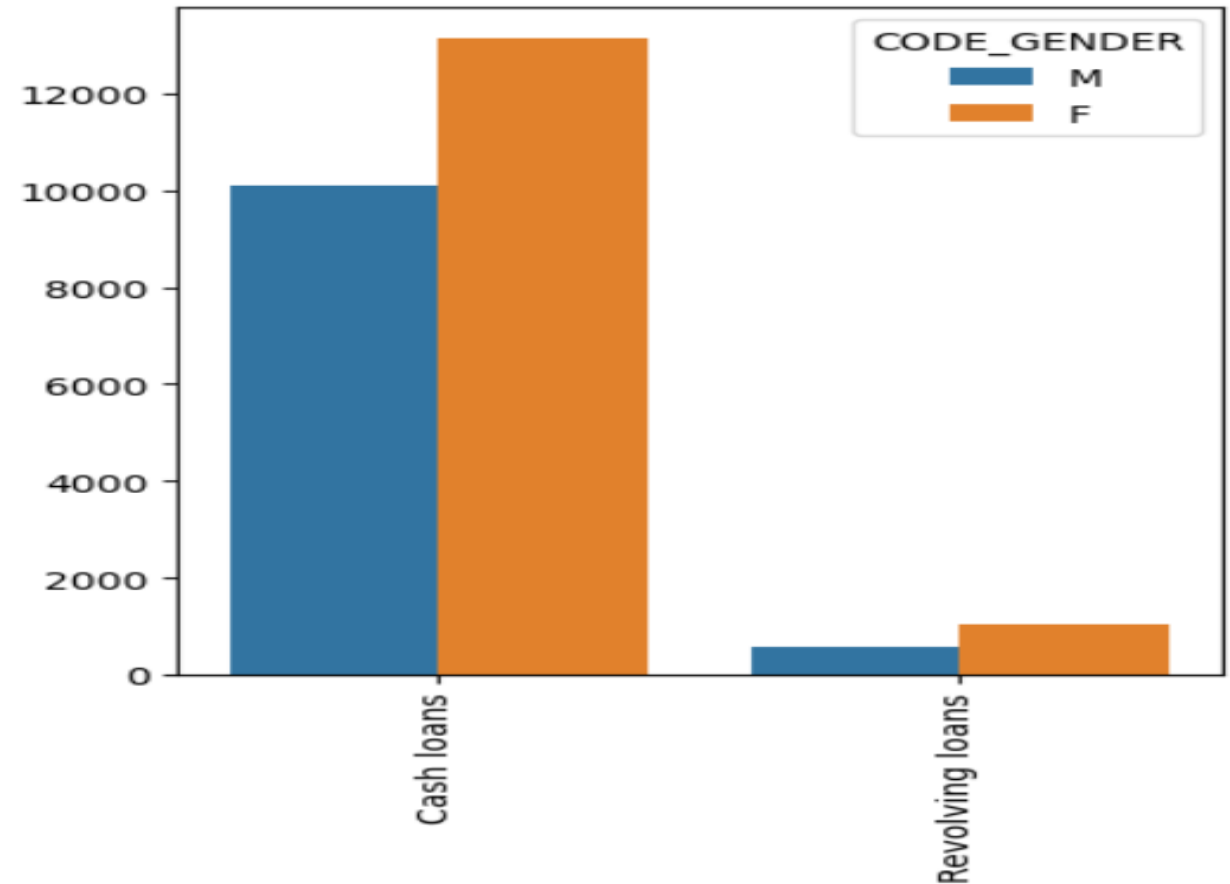
- ▶ Labours are having the most payment difficulties in repaying the loan amount.
- ▶ HR and IT staff are negligible in this comparison.

Customer with payment difficulties

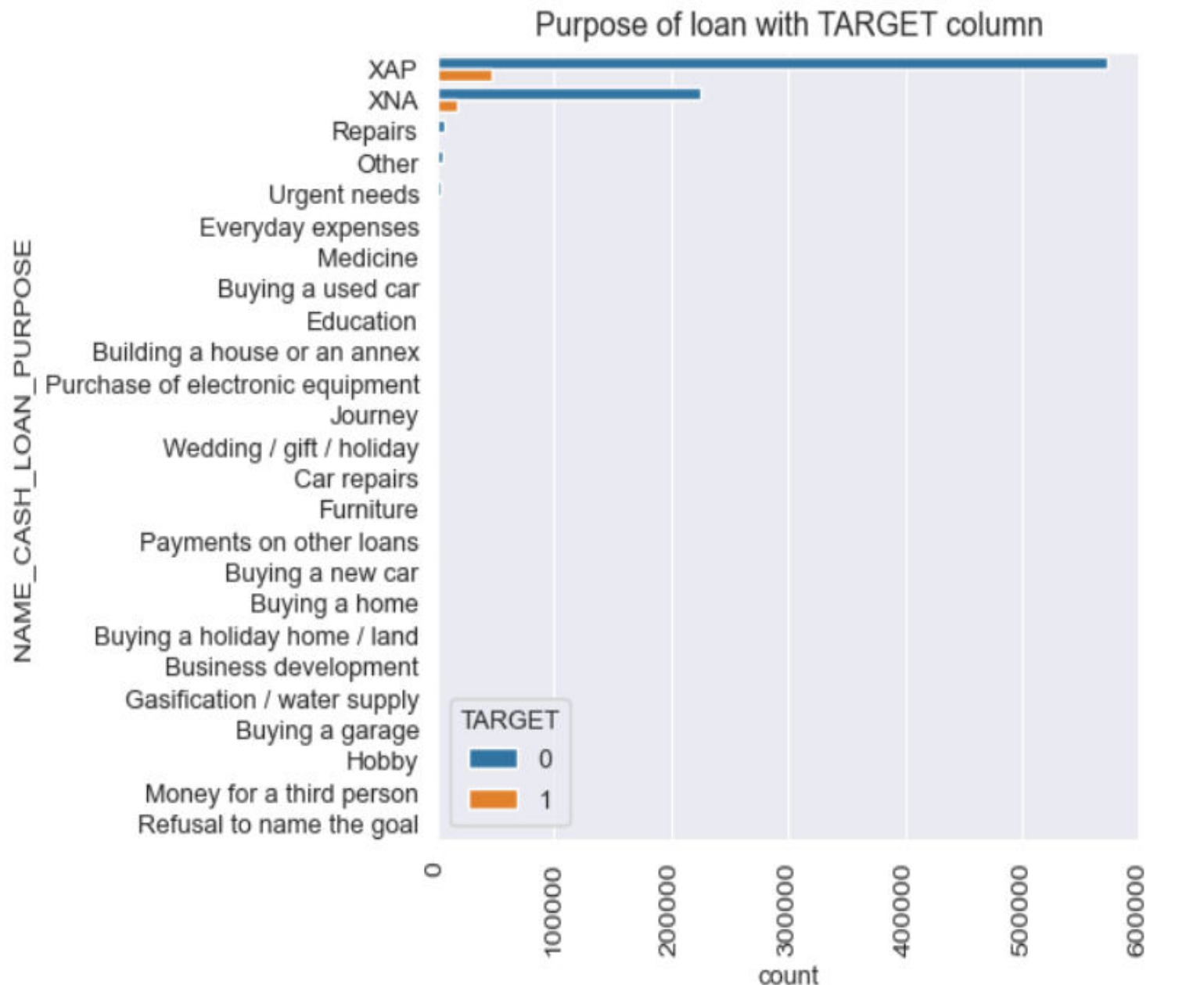


Code Gender

- ▶ Female customers are more defaulters than male.
- ▶ Females are taking more cash loans and are defaulters.

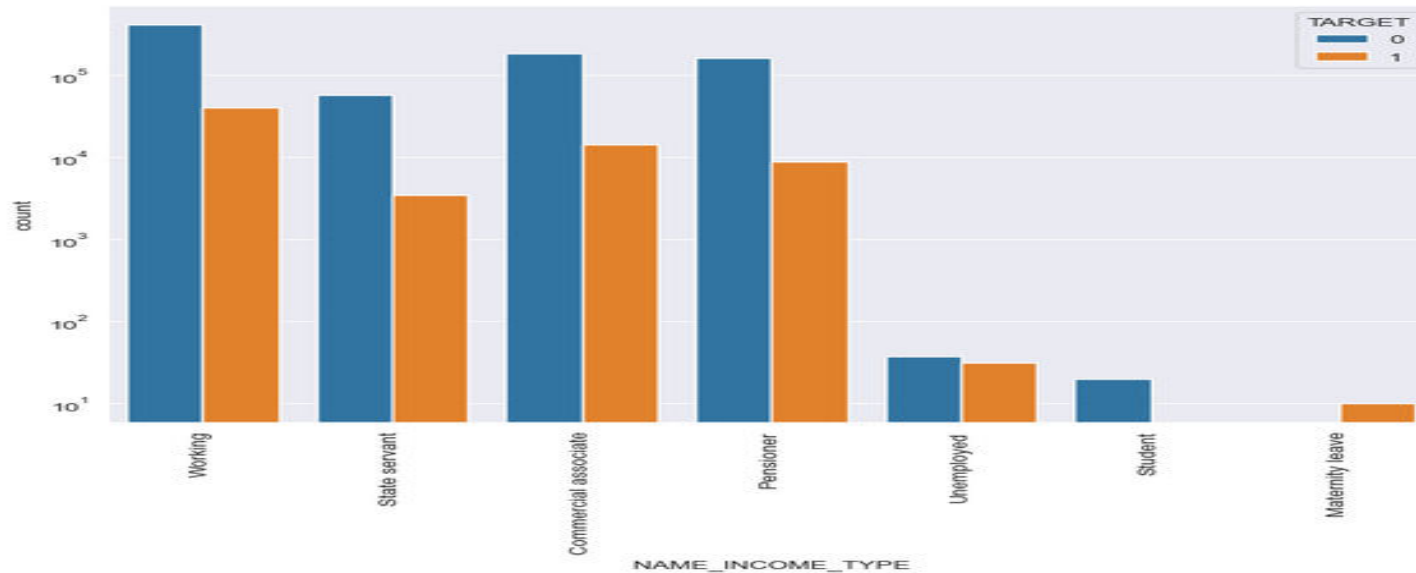
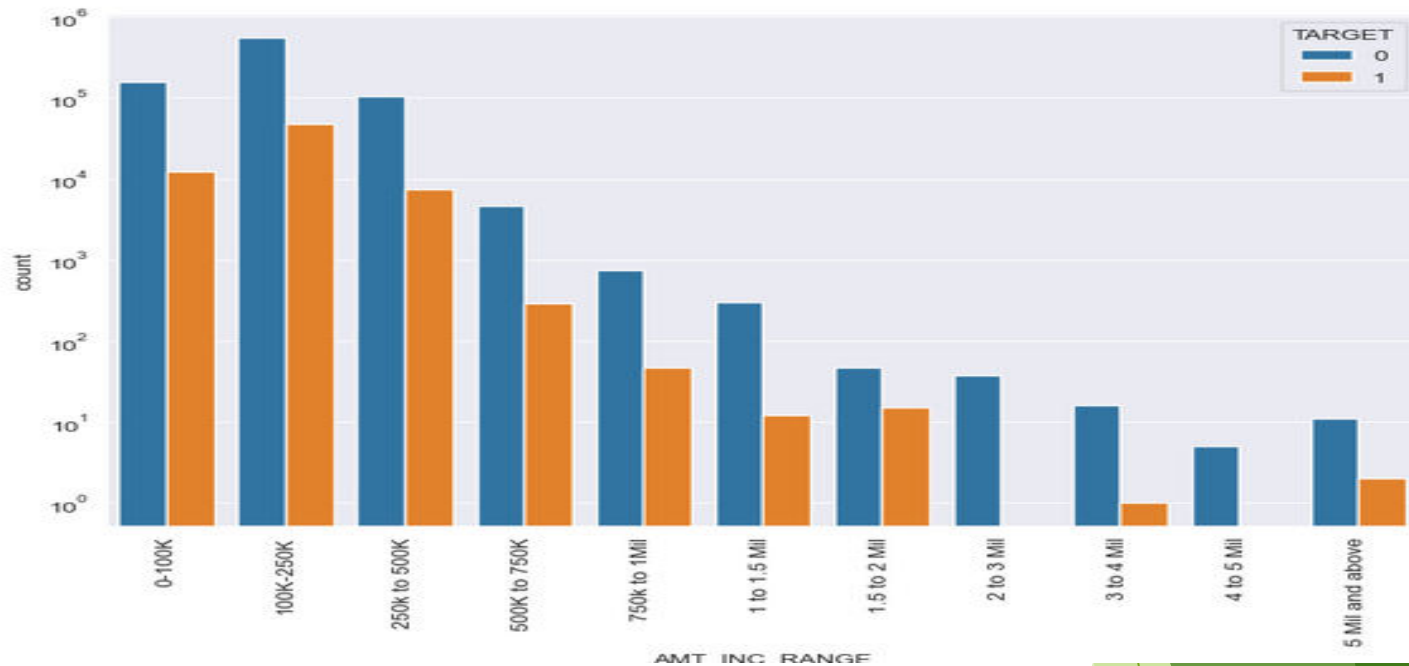


Purpose

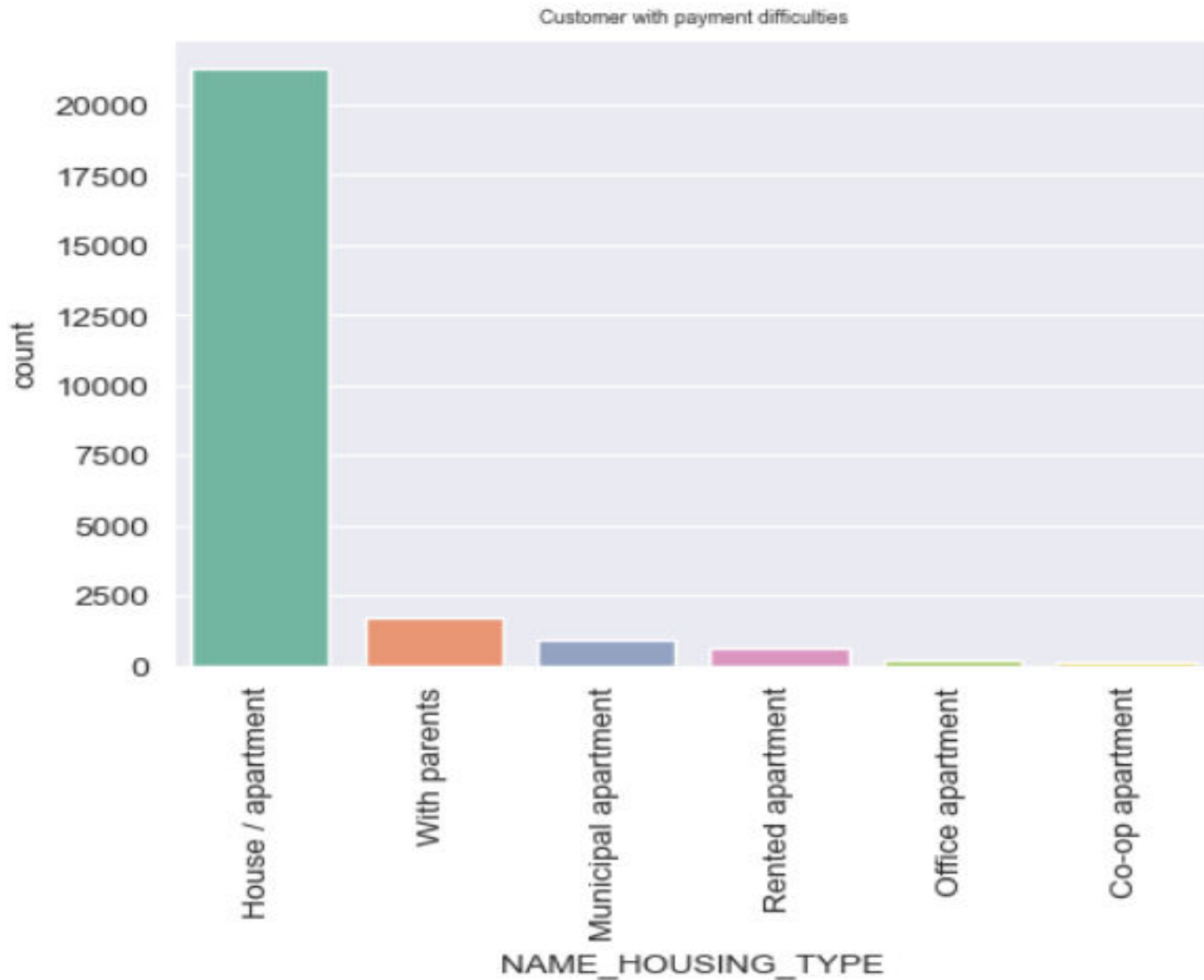


Even though the purpose of loan is not clearly mentioned there are some defaulters in this category as compared to others who have none defaulters.

Earning



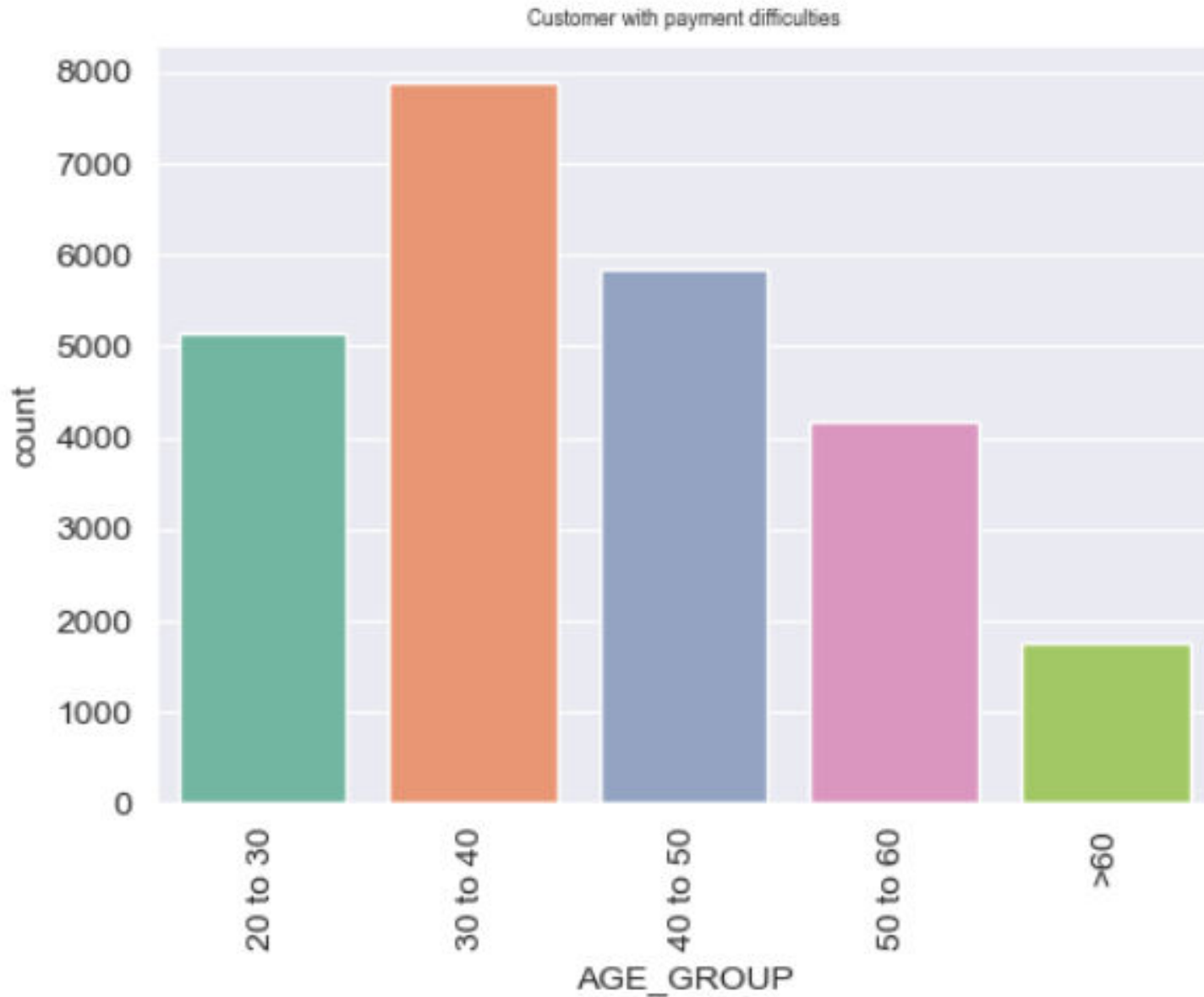
- The customers who are earning 100k-250k are more defaulters .
- The customers who are earning 4M-5M are non-defaulters.
- Students are most likely to be non-defaulters



Ownership

- Clients who own house/apartment are doing more default compared to other categories.

Age Group



- Clients of age group 30-40 are having more payment difficulties.

Conclusion

- ▶ Bank should avoid giving loans to labours as they are more likely to be defaulters.
- ▶ Bank should also check the gender of applicant as the female customers are more defaulters.
- ▶ Bank should closely check purpose of loan as people are not mentioning their purpose.
- ▶ Bank should check the income level of customers before giving the loan.



► Thank You

By- Ruchi Dham