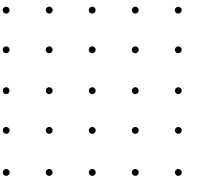# Credit Risk Analysis

-By Keshav Mishra
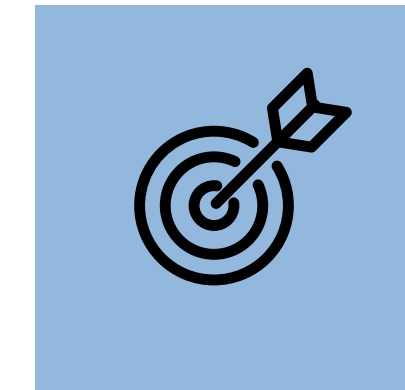
Problem Statement & Objective

Problem Statement

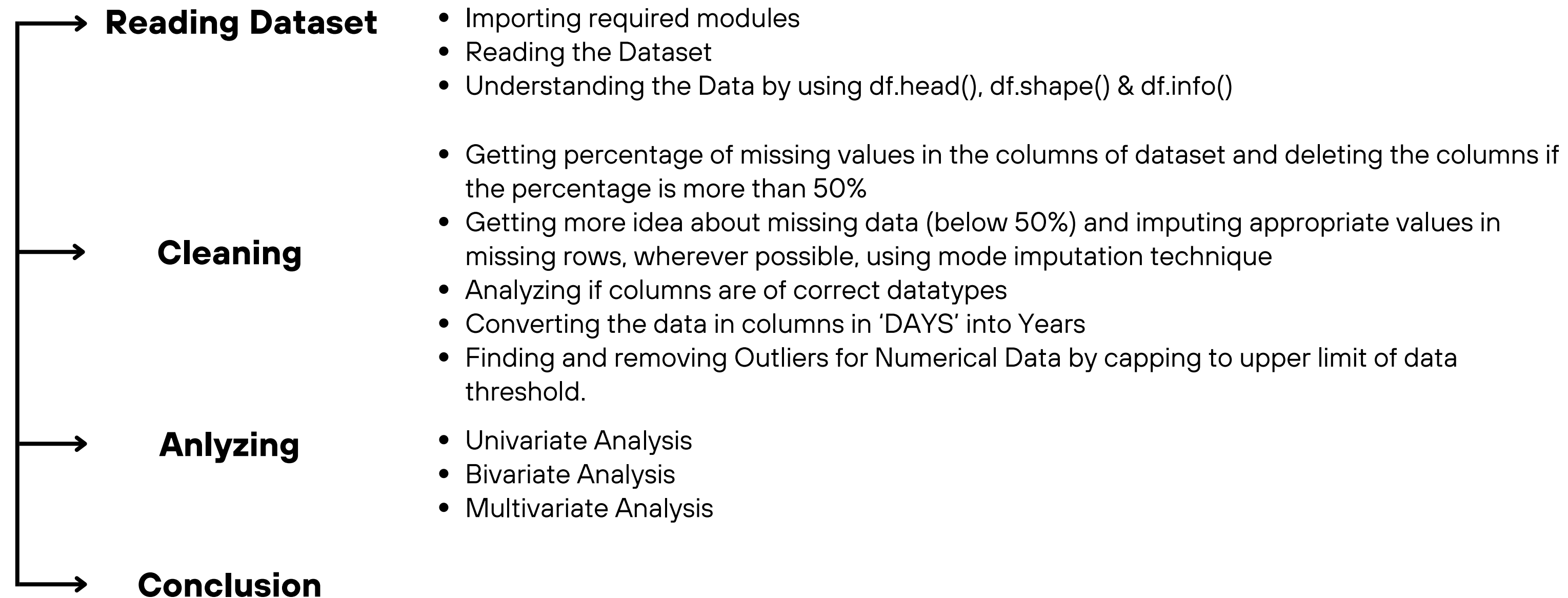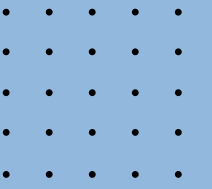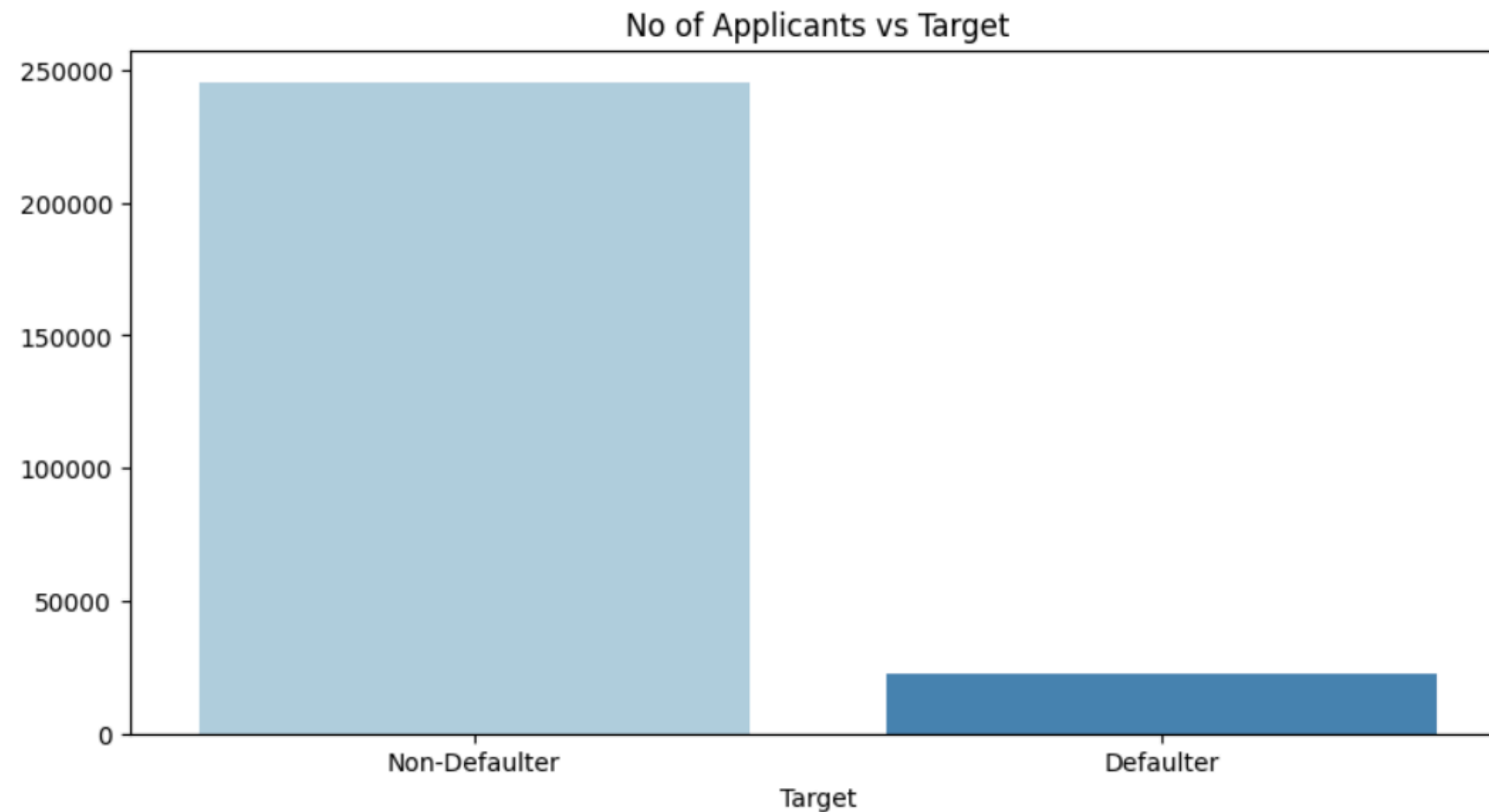To perform EDA on a loan application data set

Objective

This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Algorithm

Reading Dataset
- Importing required modules
- Reading the Dataset
- Understanding the Data by using df.head(), df.shape() & df.info()

Cleaning
- Getting percentage of missing values in the columns of dataset and deleting the columns if the percentage is more than 50%
- Getting more idea about missing data (below 50%) and imputing appropriate values in missing rows, wherever possible, using mode imputation technique
- Analyzing if columns are of correct datatypes
- Converting the data in columns in 'DAYS' into Years
- Finding and removing Outliers for Numerical Data by capping to upper limit of data threshold.

Anlyzing
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
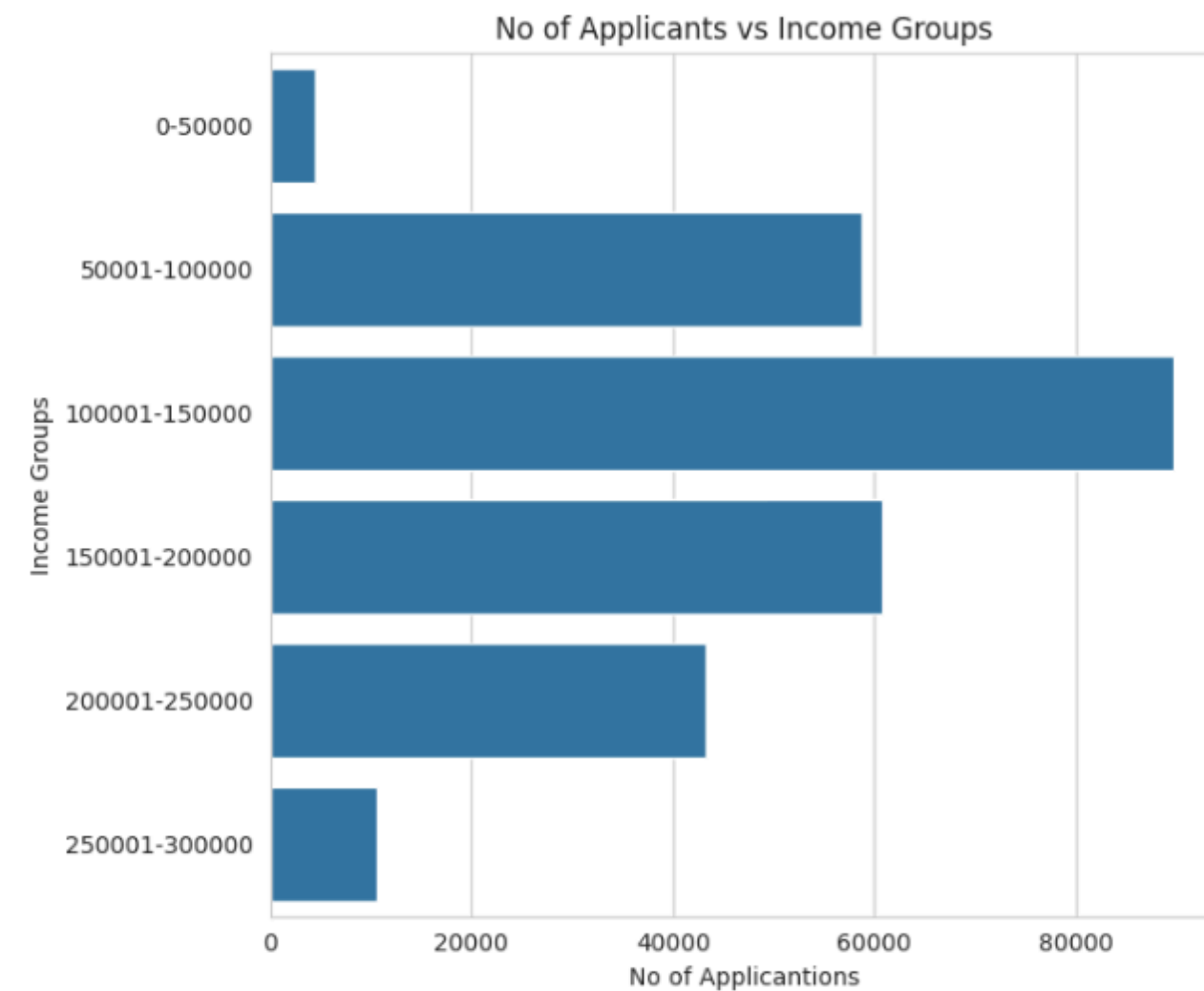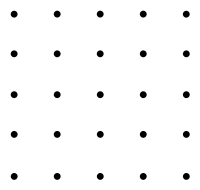
Conclusion

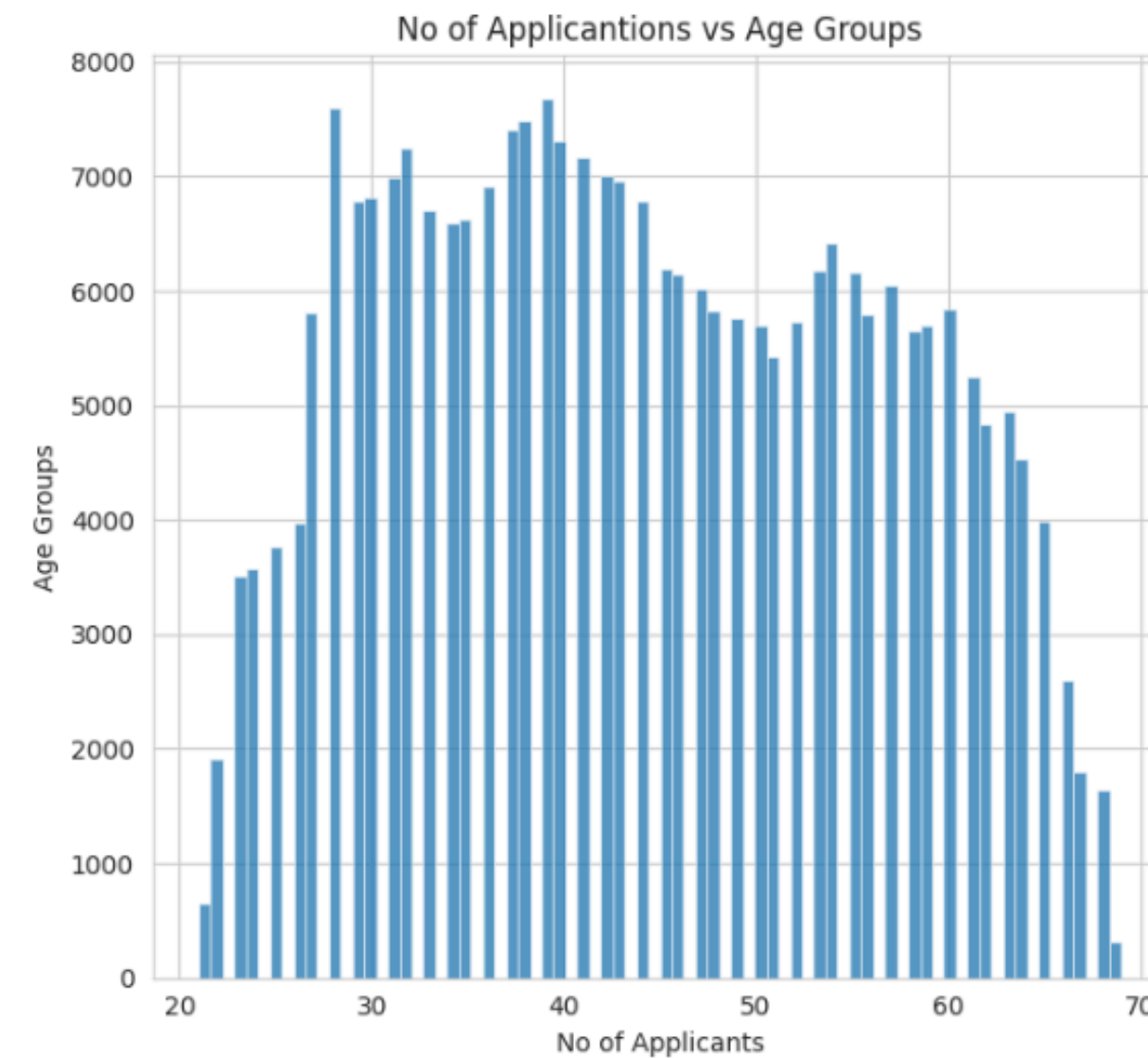No of Applicants vs Target

Defaulters v/s Non Defaulters

Data Analysis

-For loan applicants the repayers (non defaulter) are significantly higher than the defaulters defaulters and non defaulters being women more likely than men.
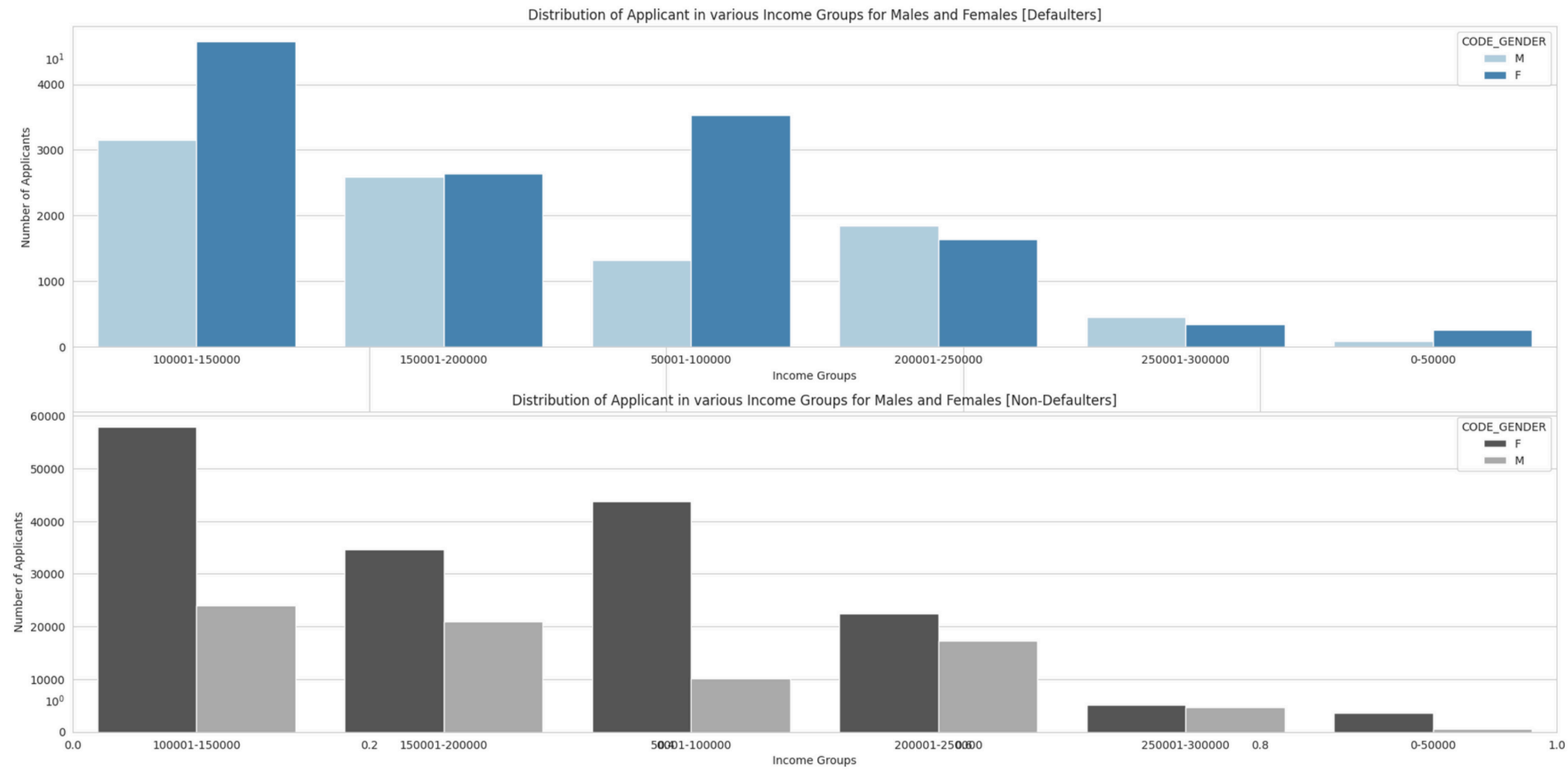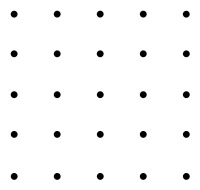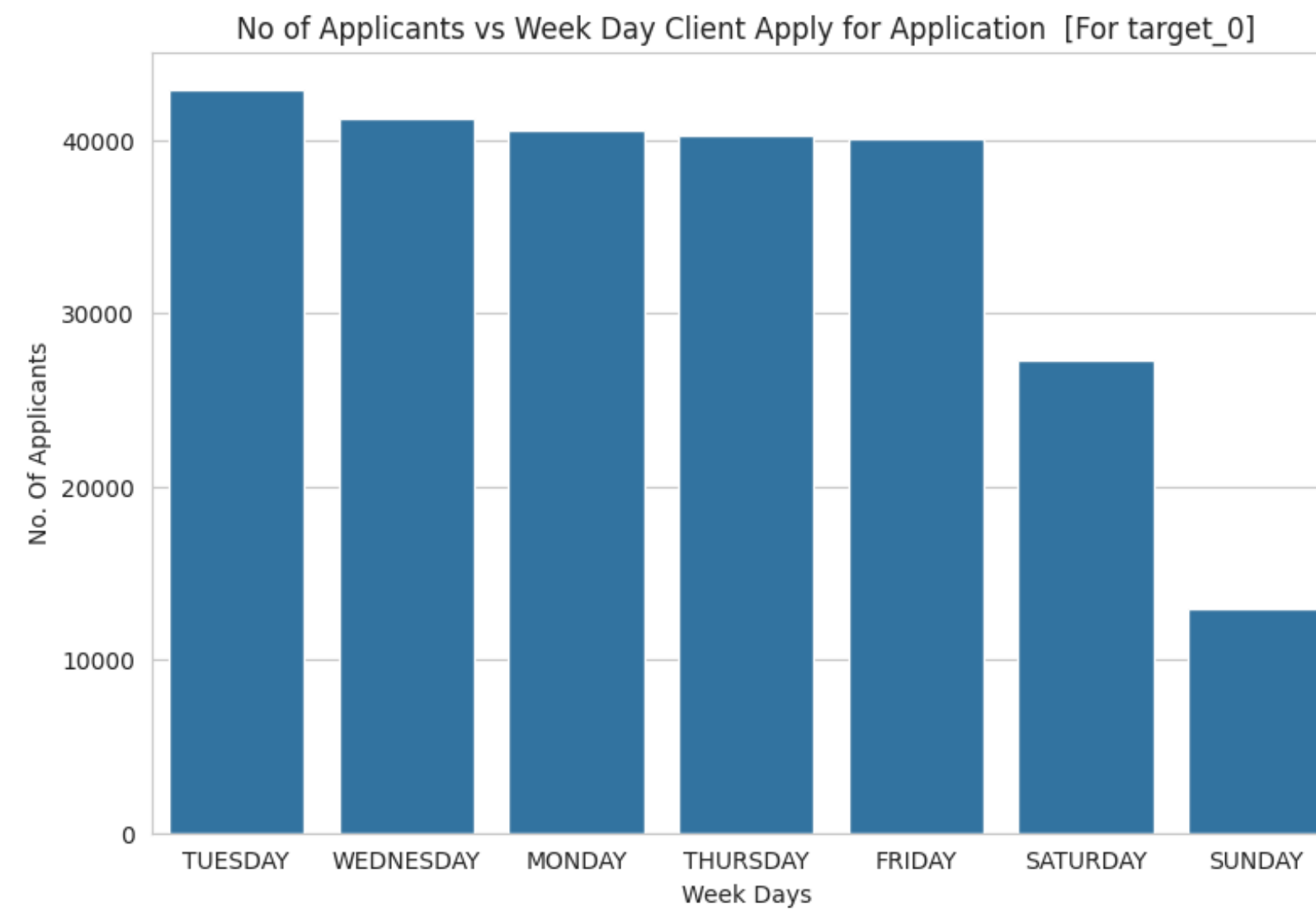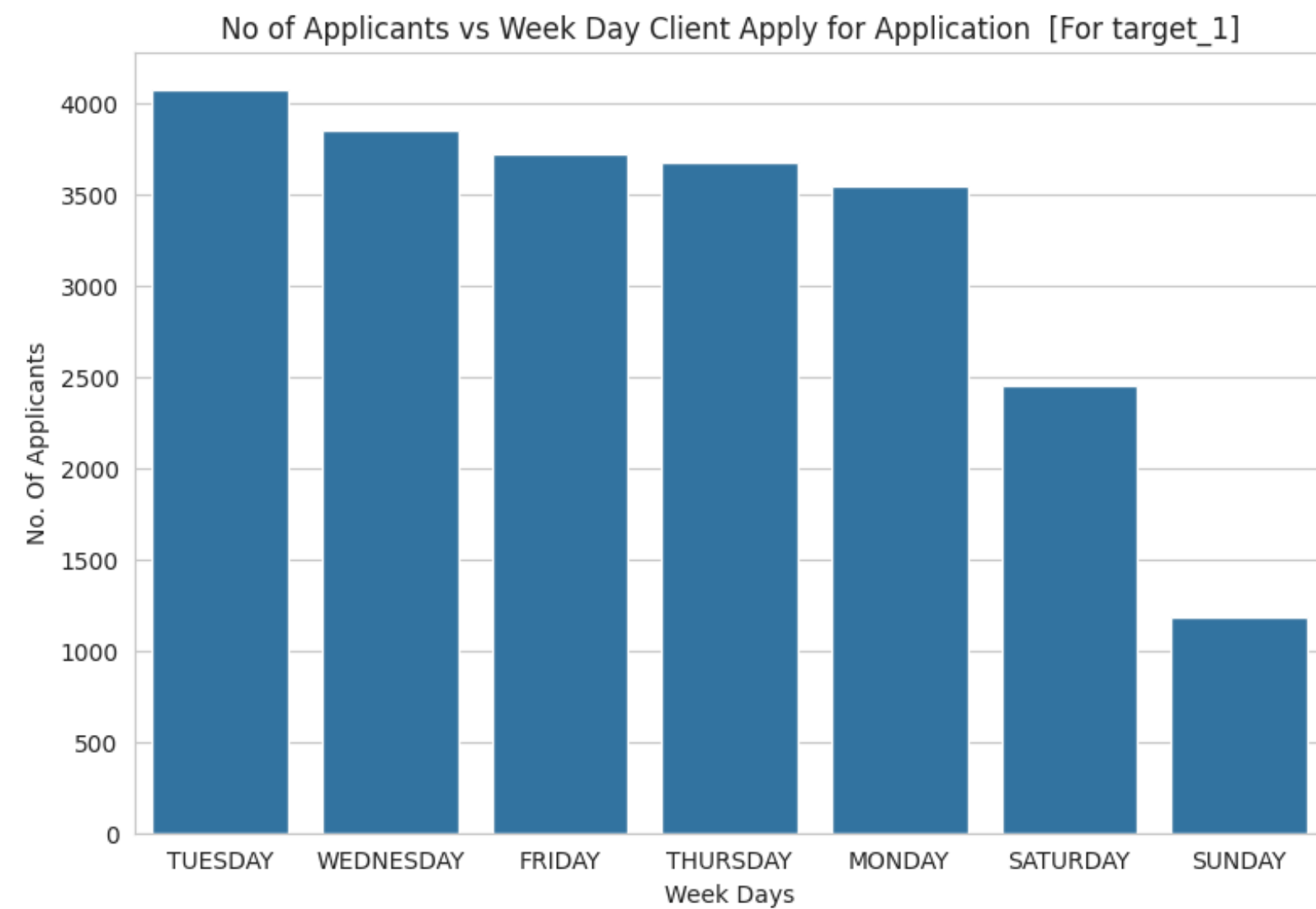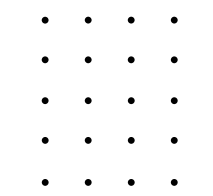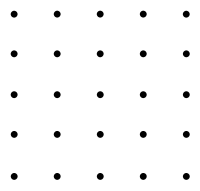- If we give out loans we have a higher chances of getting out money back.
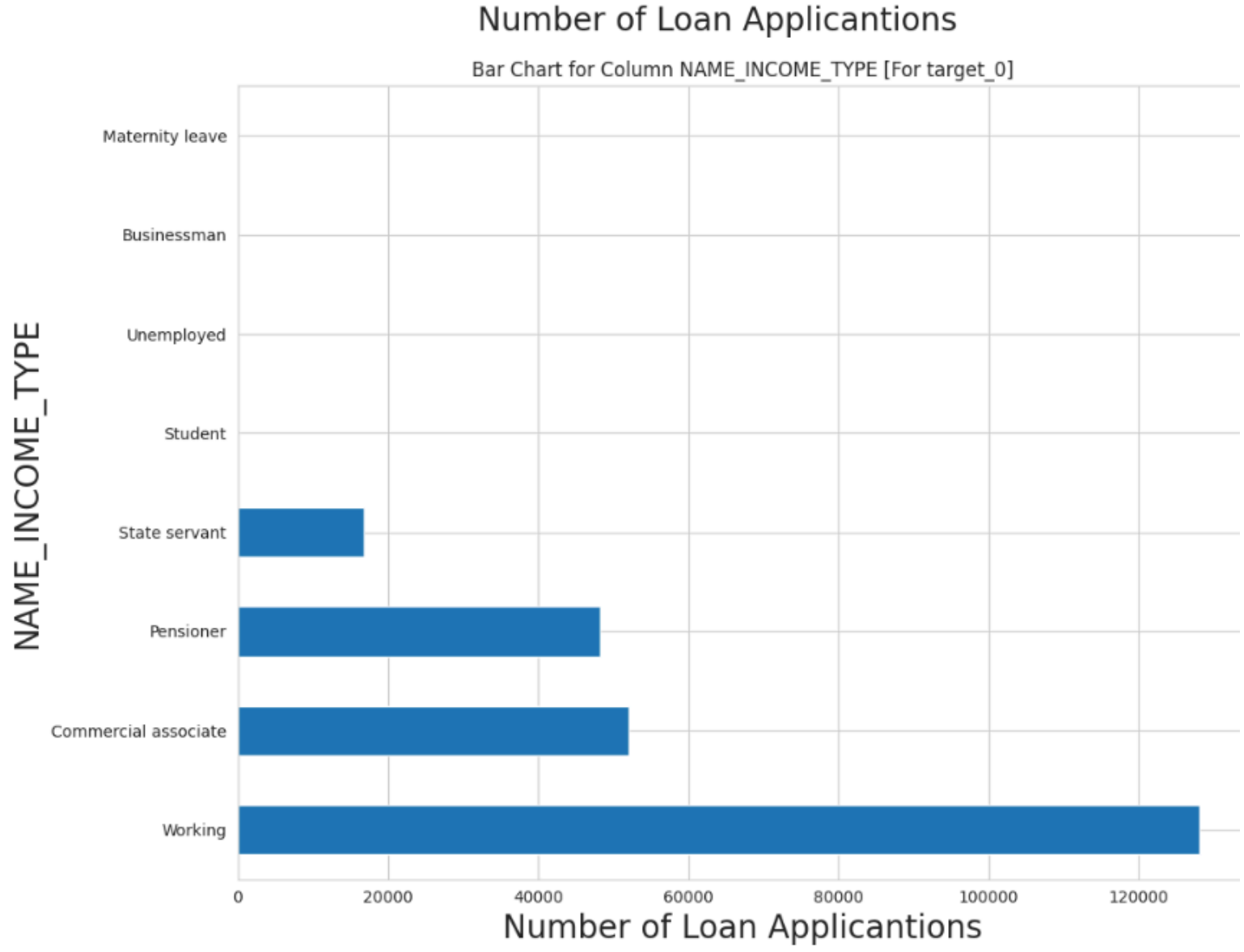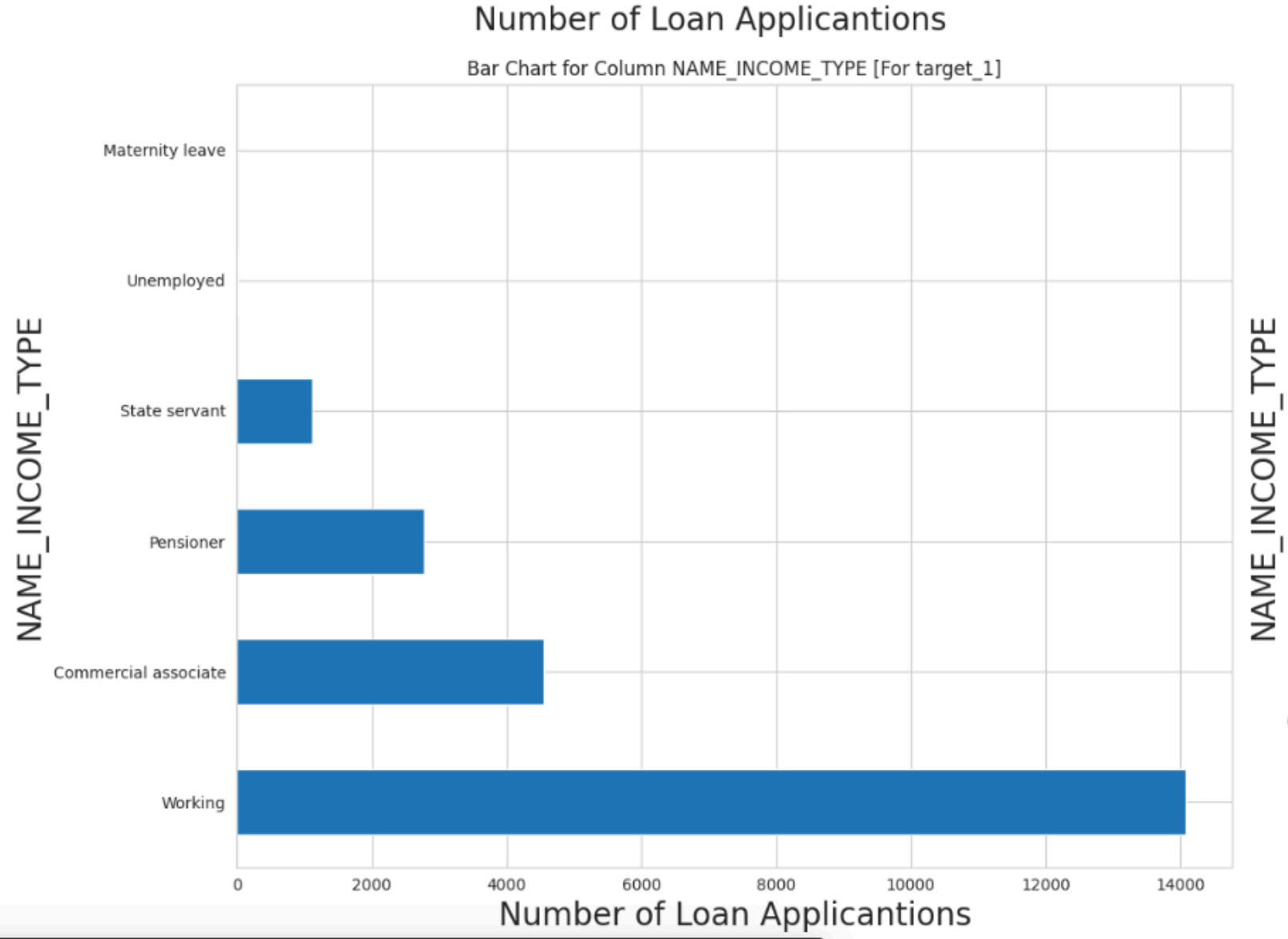
Most applicants lie in Income Group of 1L to 1.5L
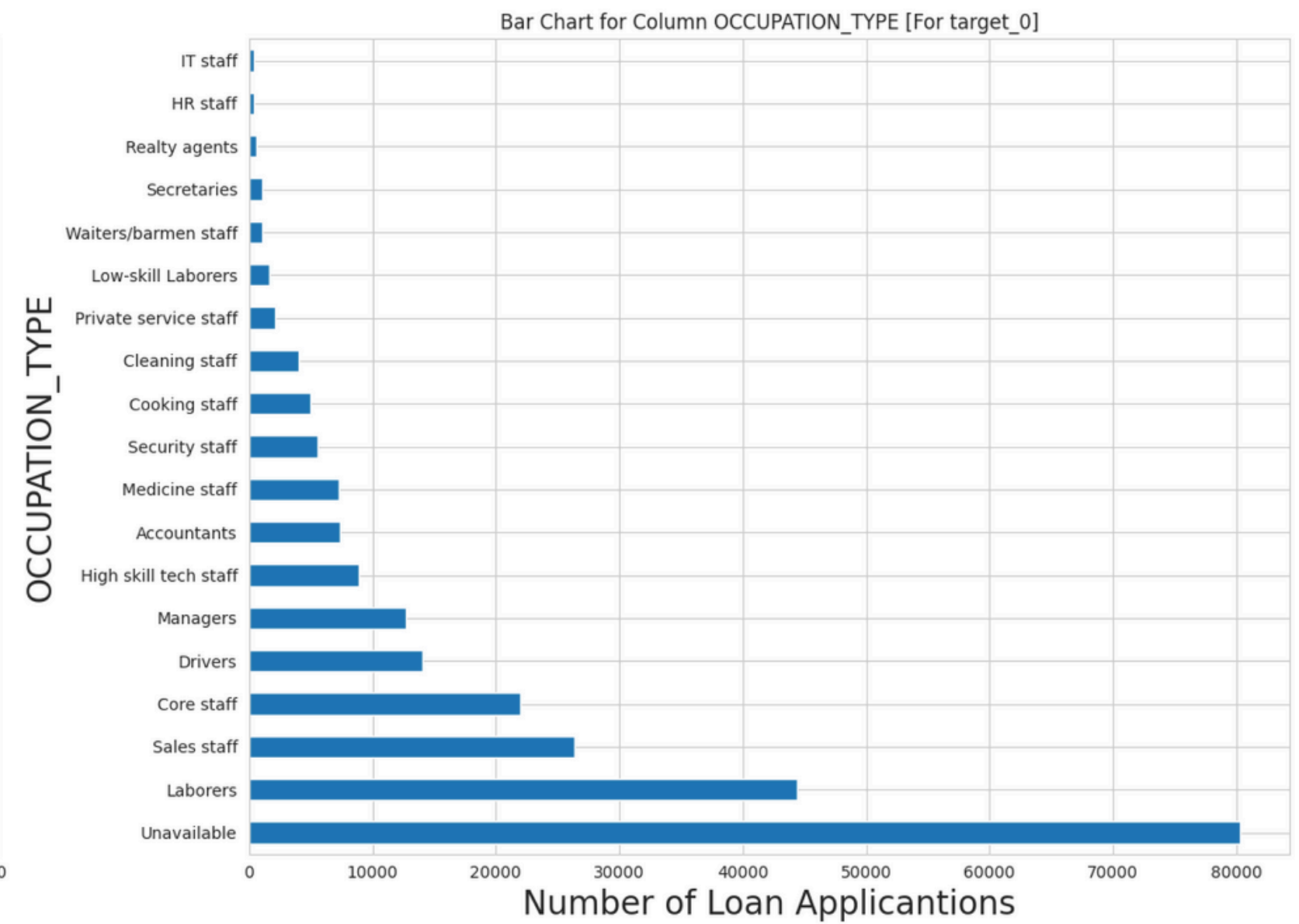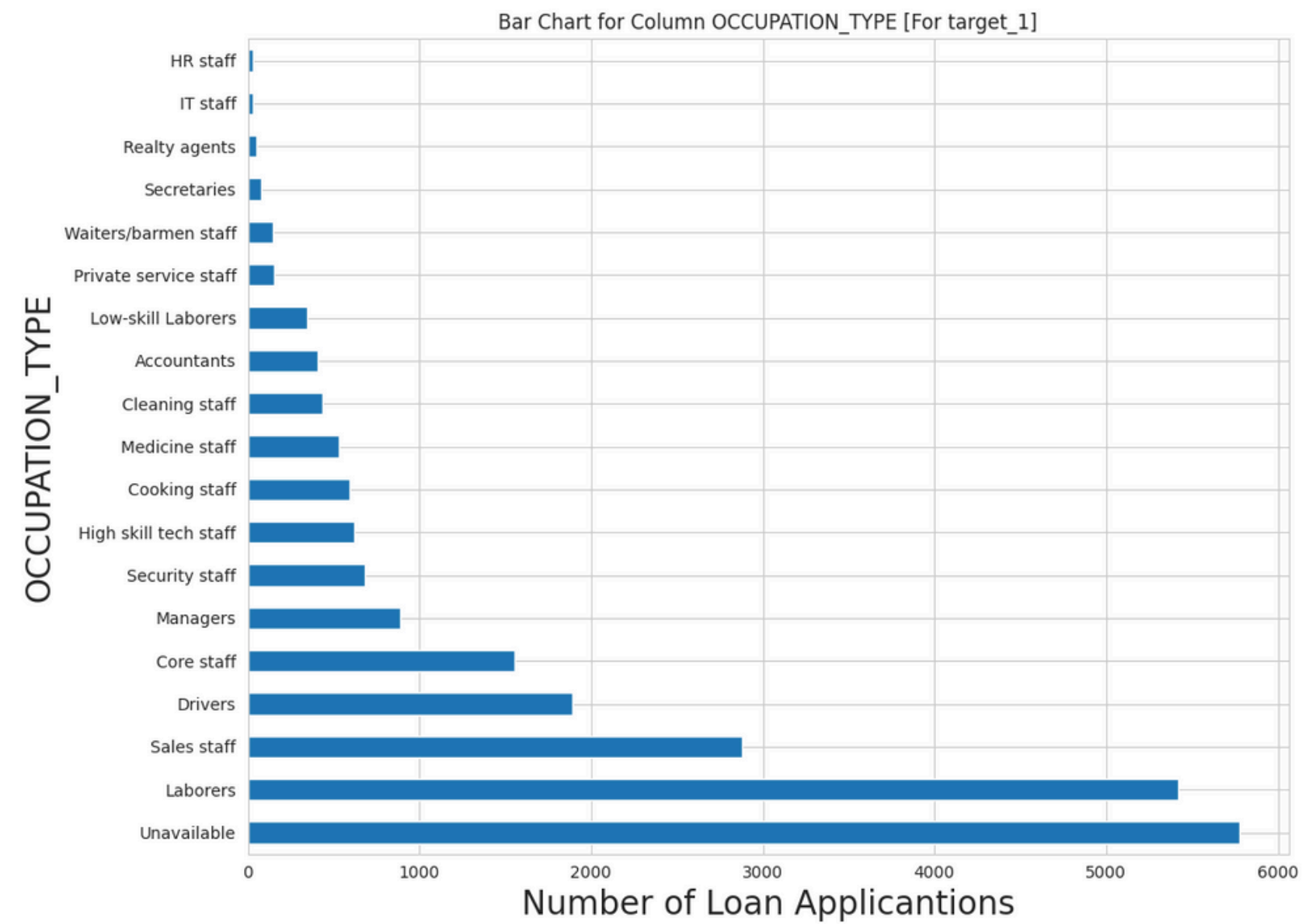

Most applicants lie in Age Group of 30-40

Distribution of Applicant in various Income Groups for Males and Females [Defaulters]

Distribution of Applicant in various Income Groups for Males and Females [Non-Defaulters]

No of Applicants vs Week Day Client Apply for Application [For target_1]

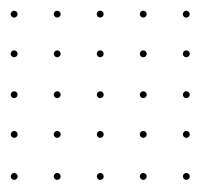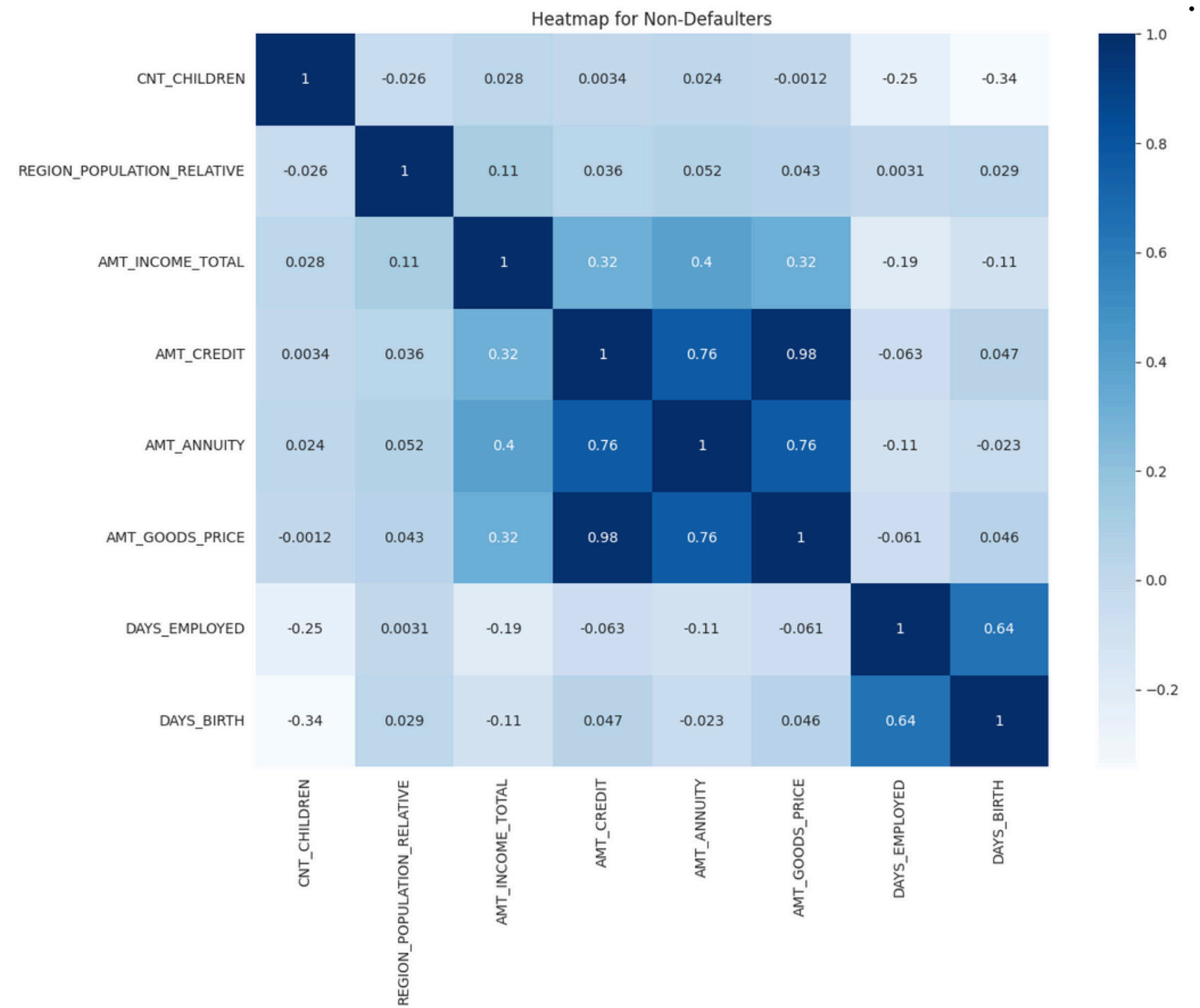No of Applicants vs Week Day Client Apply for Application [For target_0]
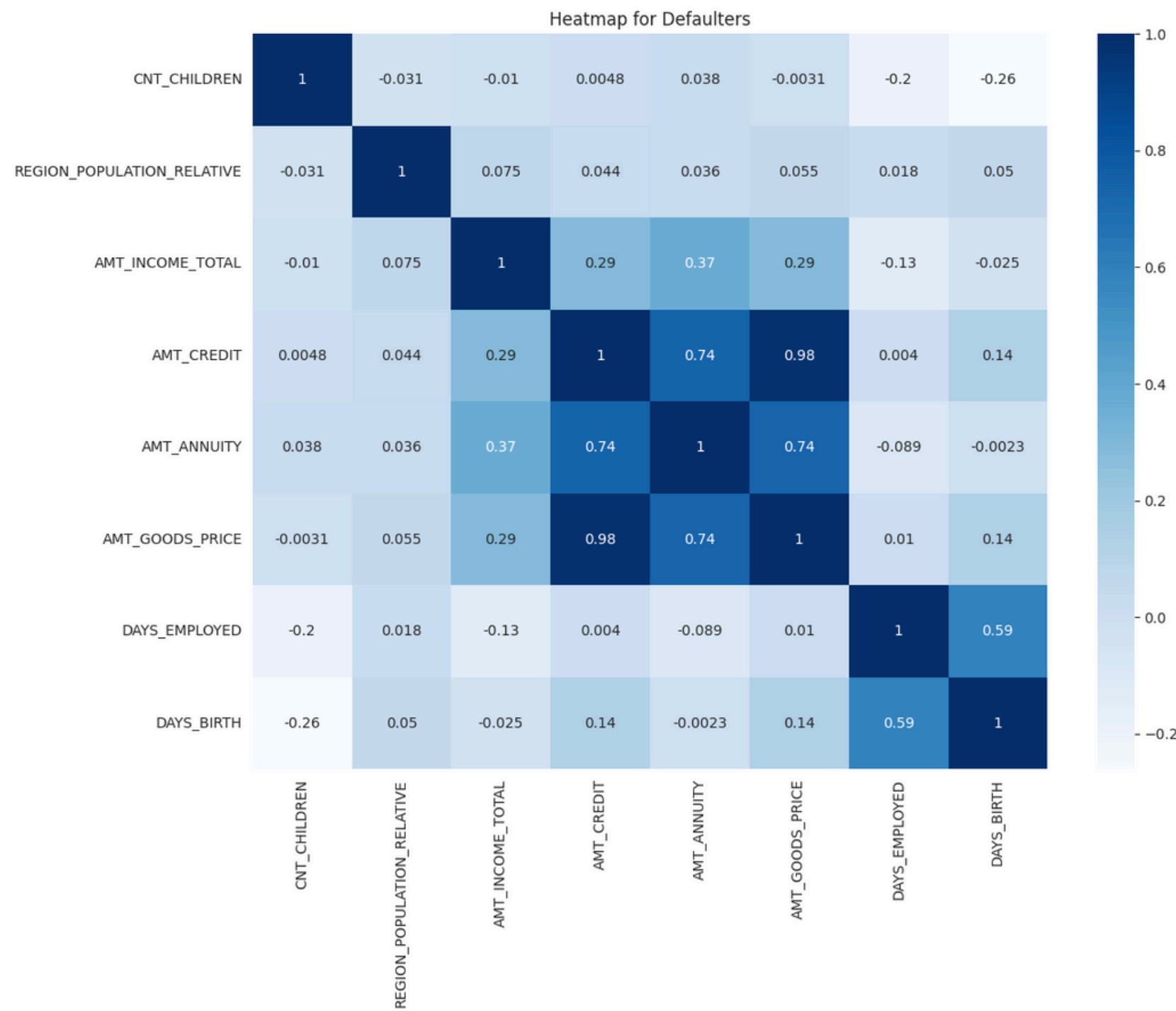
Number of Loan Application as per Occupation for Defaulters and Non Defaulters

Bar Chart for Column OCCUPATION_TYPE [For target_1]

Bar Chart for Column OCCUPATION_TYPE [For target_0]

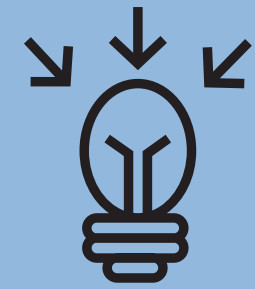Heatmap for Defaulters

Heatmap for Non-Defaulters

Strong correlation between the Goods Price and Loan Amount that was credited

Distribution of credited loan amount with Education Status for Loan Defaulters
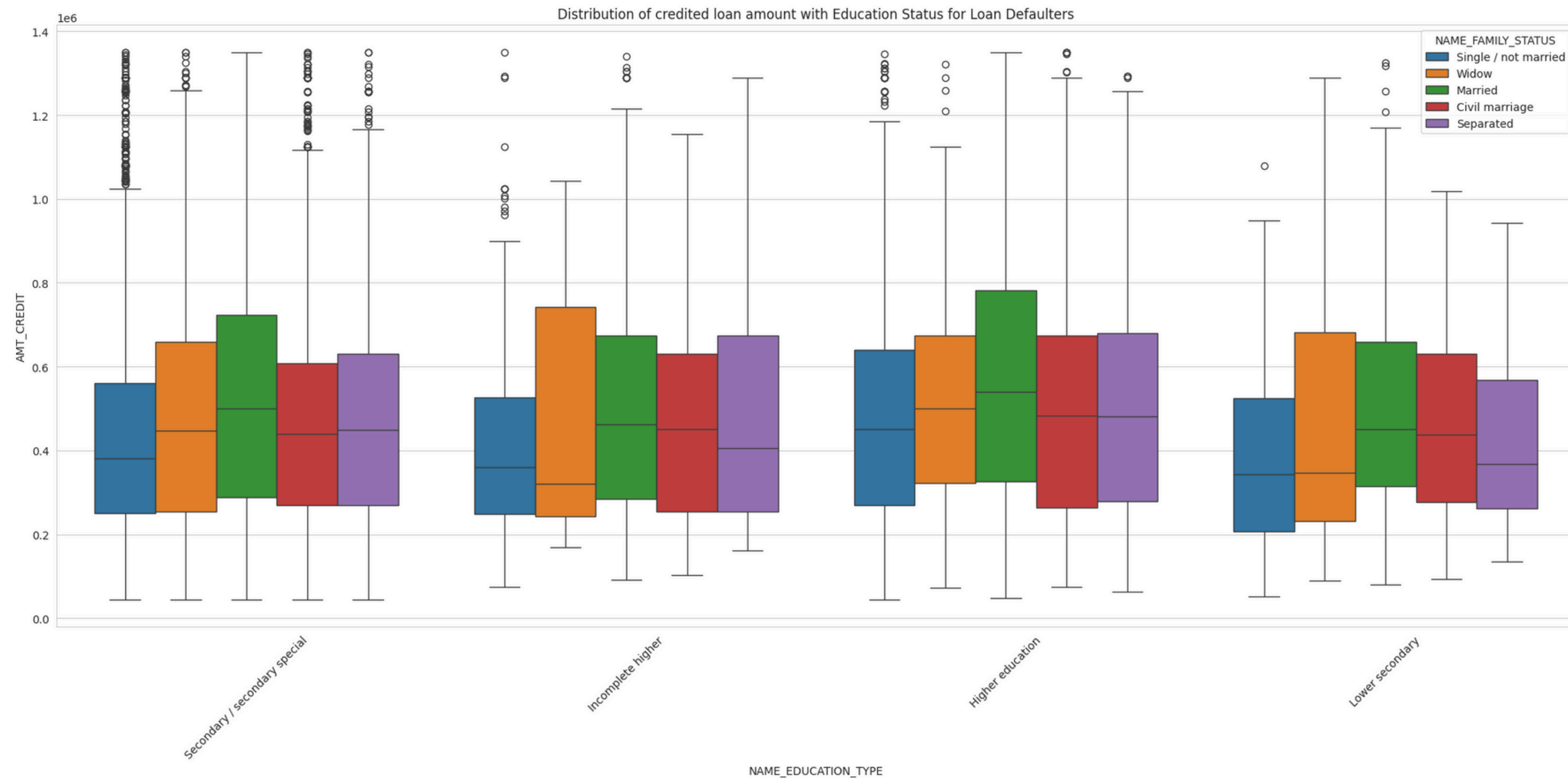
Married people with higher education have defaulted loans for higher distribution of credited loan amounts.
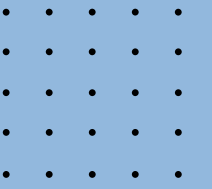
- Applicants having an academic degree have defaulted less number of loans compared to applicants from other educational backgrounds
- Married people with higher Secondary education background have defaulted loans for higher distribution of credited loan amounts.
- Number of defaulters are higher in females than males.
- Bank is expected to receive higher number of loan applications on Tuesdays than other days

Conclusion