OGTIP PROJECT 2

EXPLORATORY DATA ANALYSIS USING PYTHON

Presented by:
DAEHYUN KO

Introduction

Exploratory Data Analysis was used in this project to analyze the indicators that the bank has to consider to reduce the chance of loan payment default from their clients.

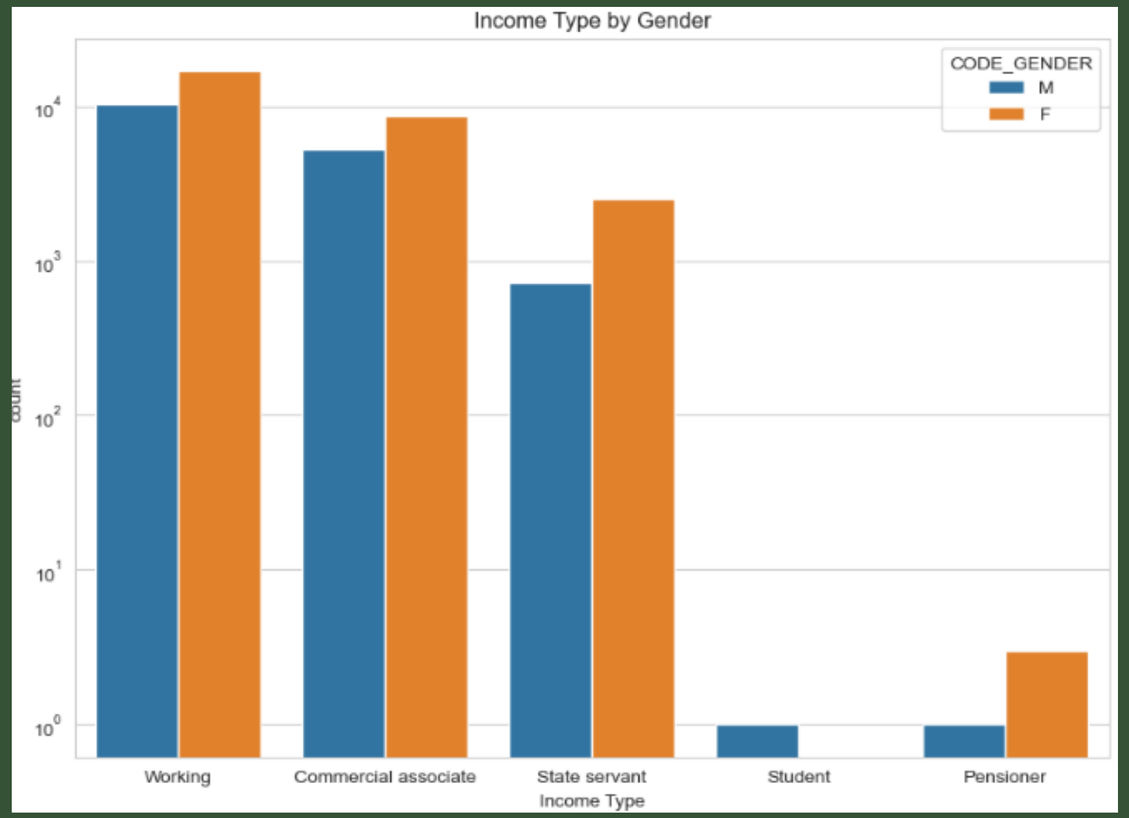
Two main datasets were analysed in Python

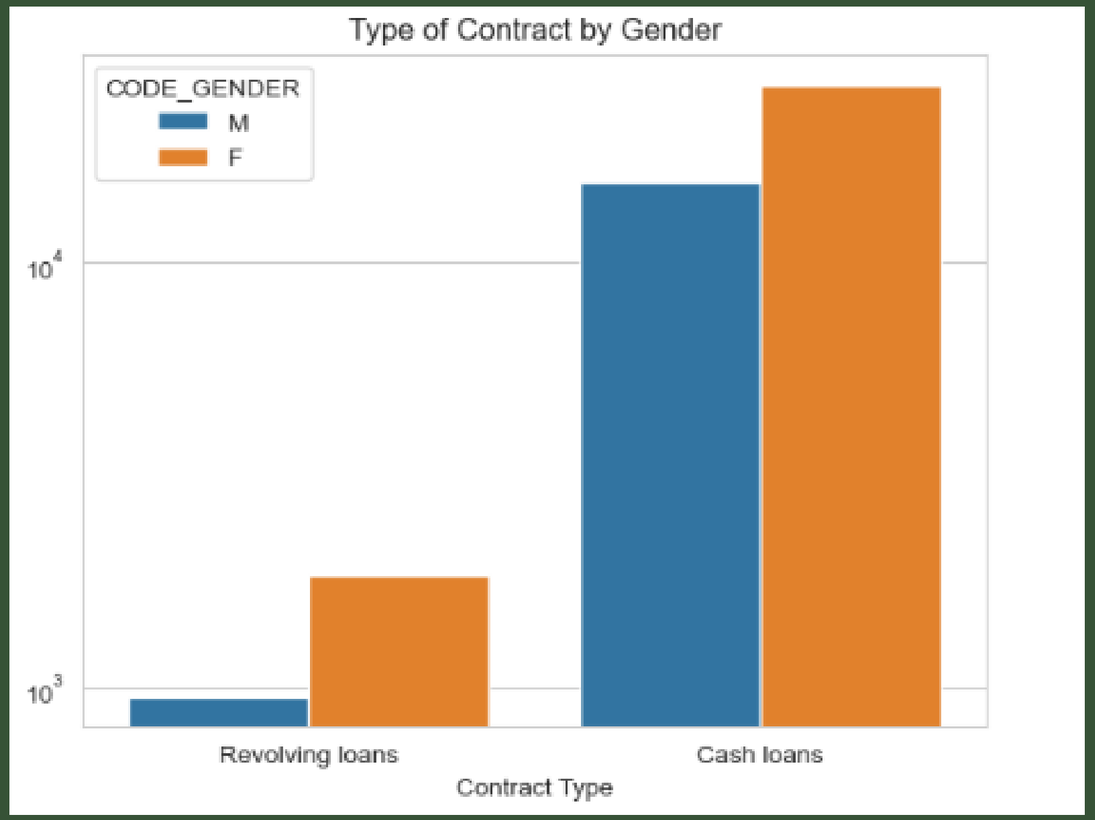
1. 'application_data.csv' contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.

2. 'previous_application.csv' contains information about the client’s previous loan data. It contains the data whether the previous application had been Approved,

Data Visualisations

Visualisation of Dataset



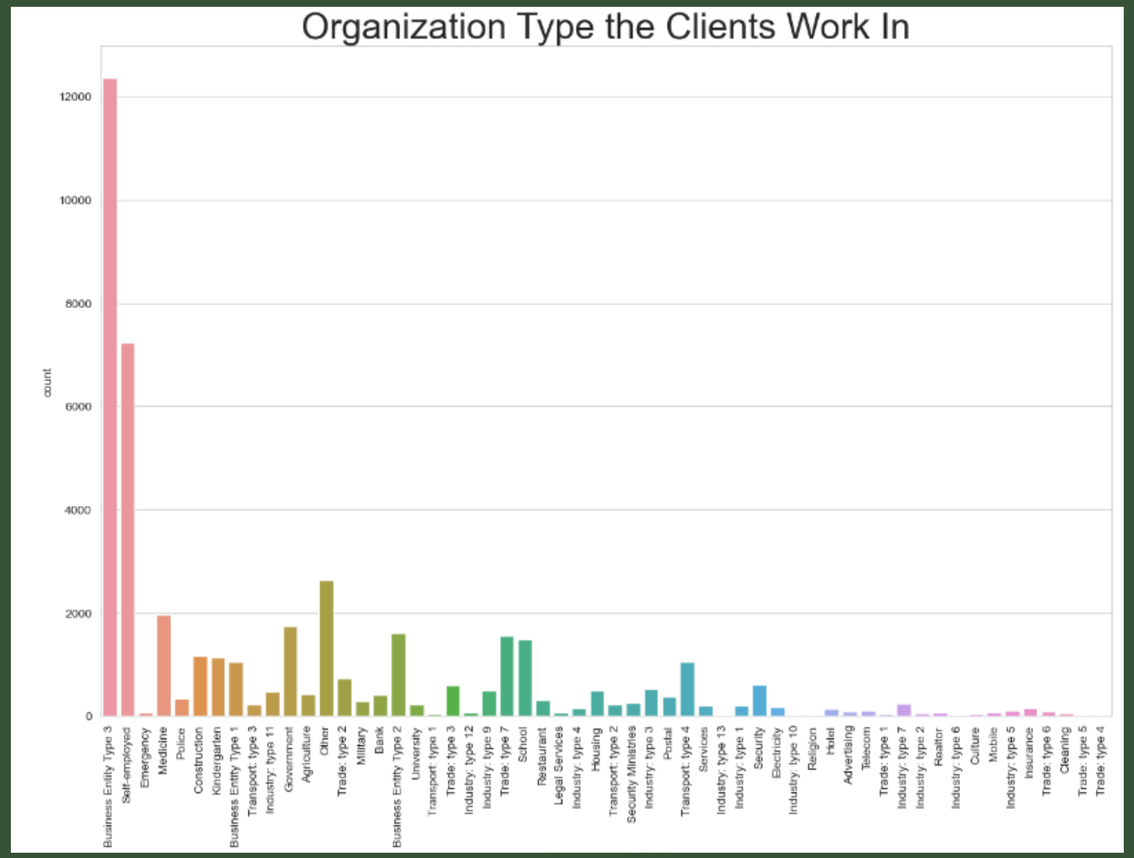


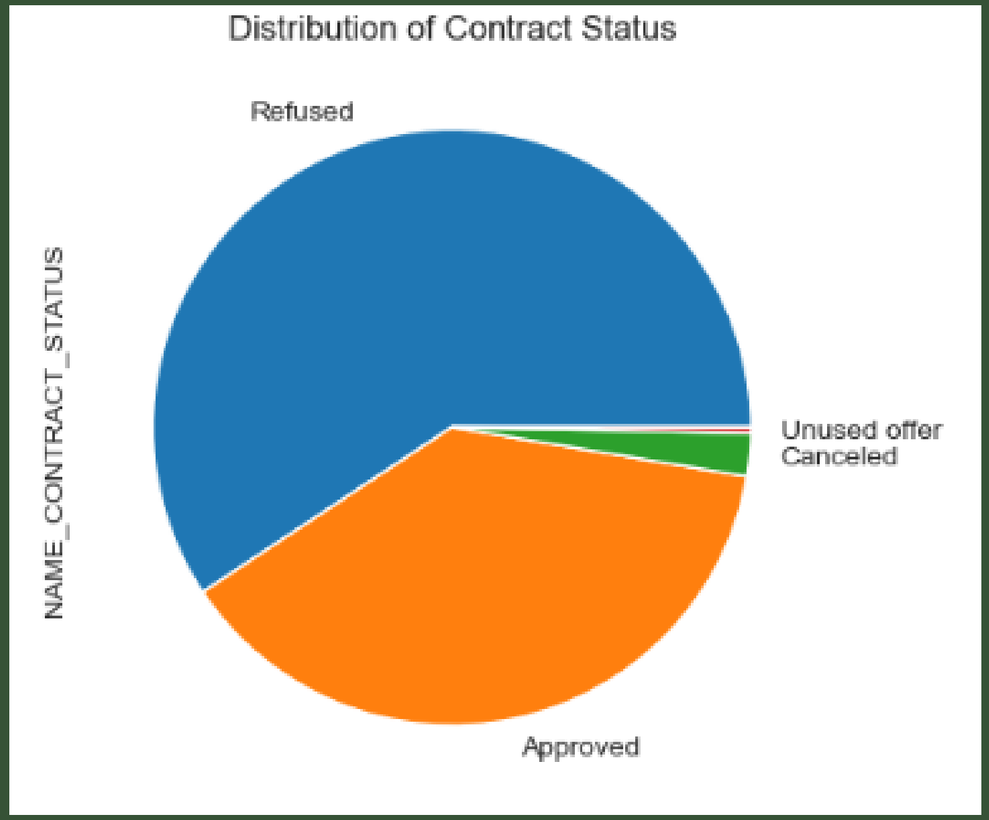
Inferences

Most loans are taken by people in the labor force

The majority of type of contract is cash loans

Visualisation of Dataset

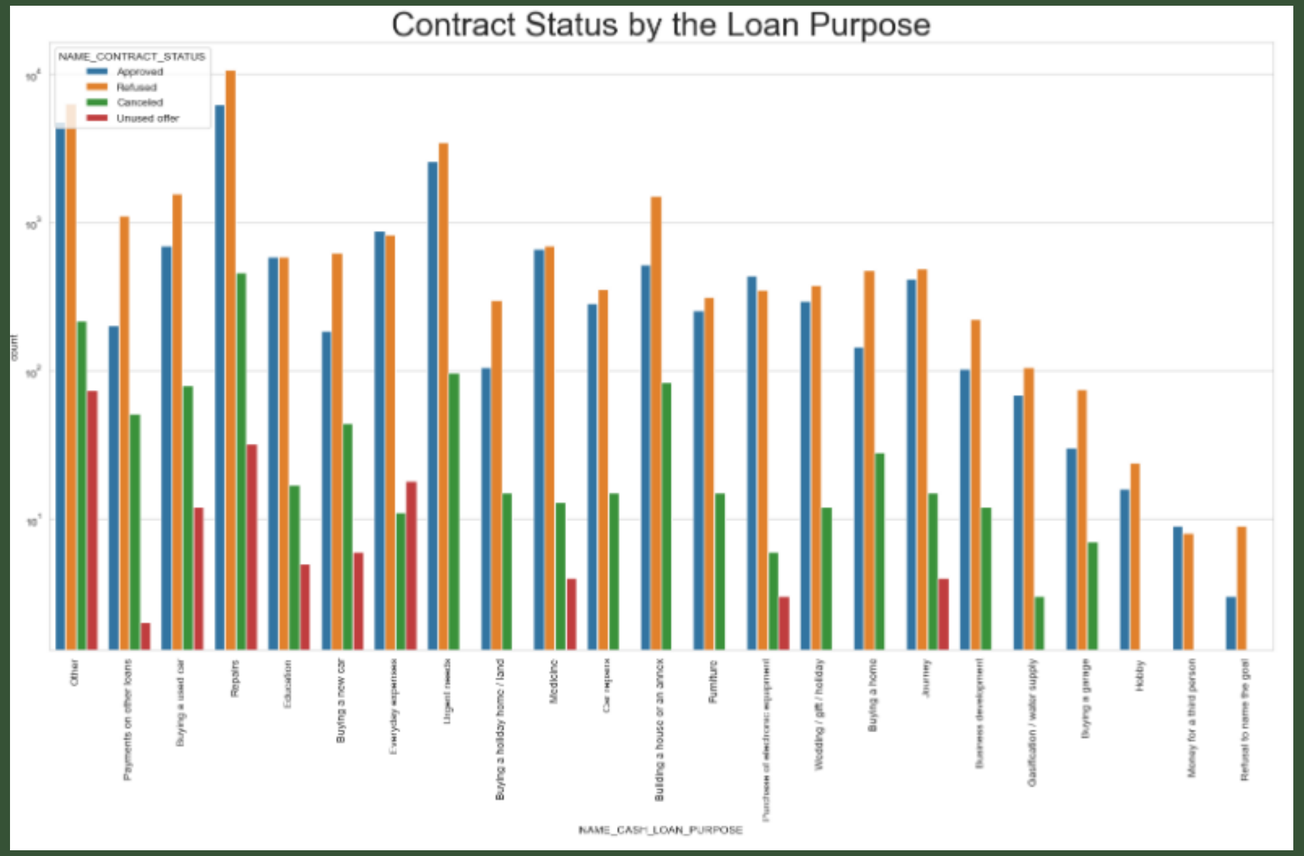




Inferences

The major organization type the client works in is Business Entity Type 3 followed by self employed
Most Applications were refused by the bank and around 40% were approved

Visualisation of Dataset



Inferences

Repairs is the most common loan purpose

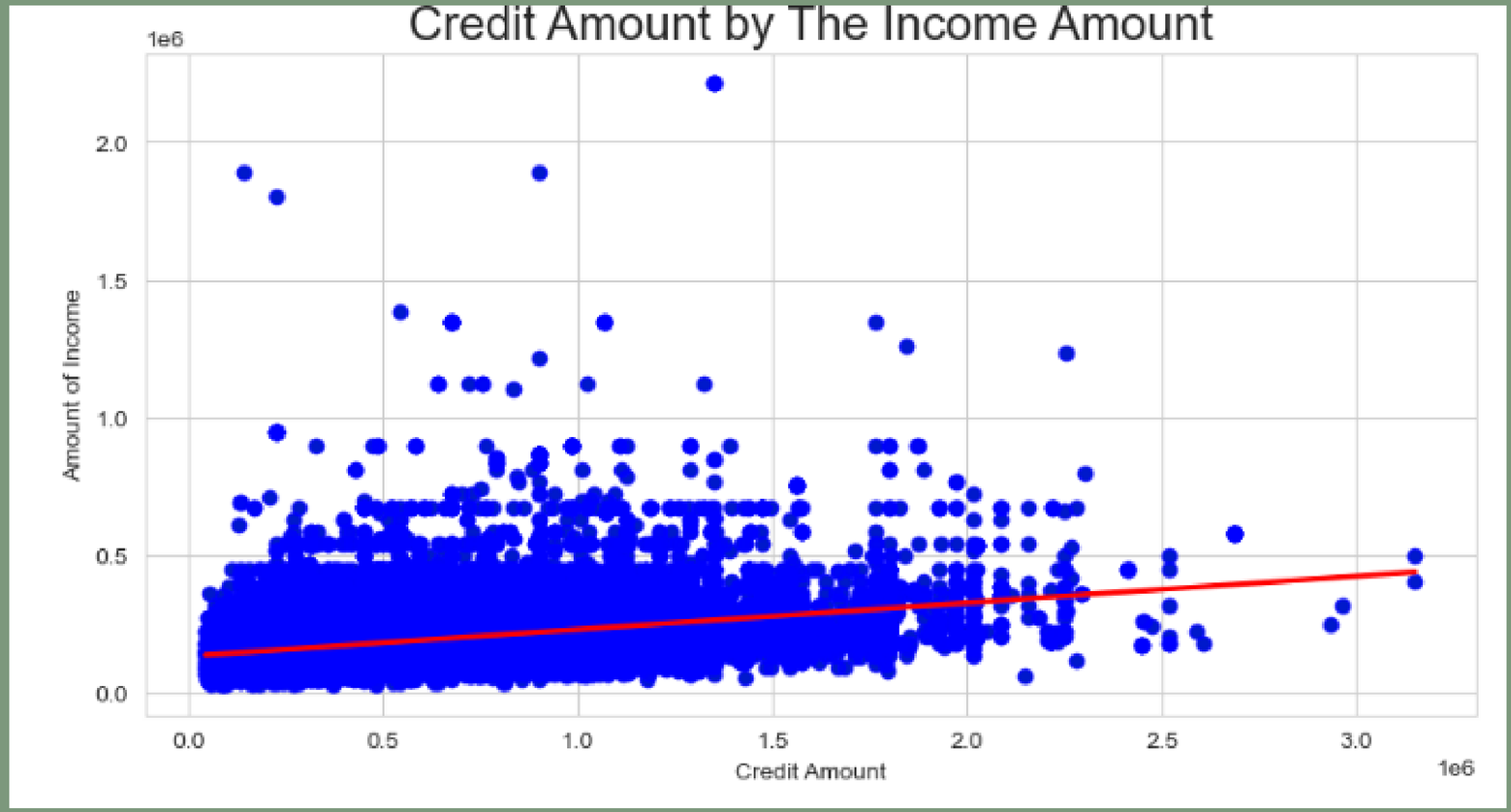
Loans for everyday use and education is most likely to be refused by the bank

"Payment for other loans" has the highest refusal rate compared to approval

Bivariate/Multivariate Analysis

Credit Amount to Income Amount

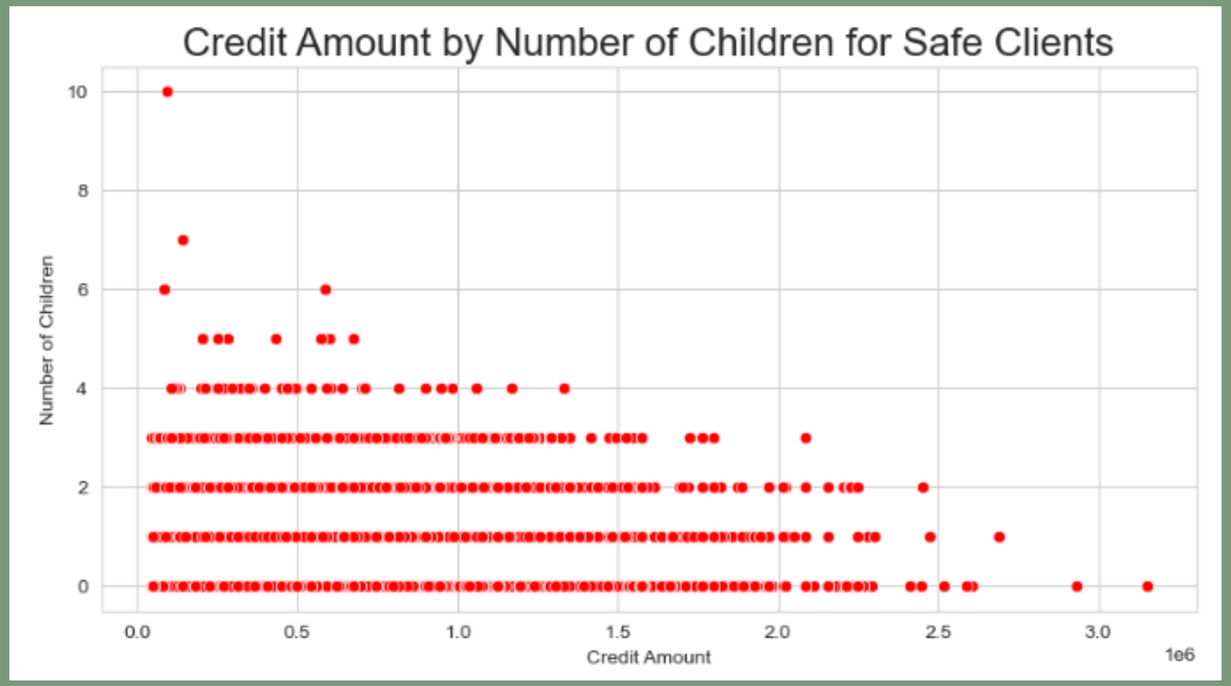
There is noticeable trend where the credit amount is positively correlated with the income amount.

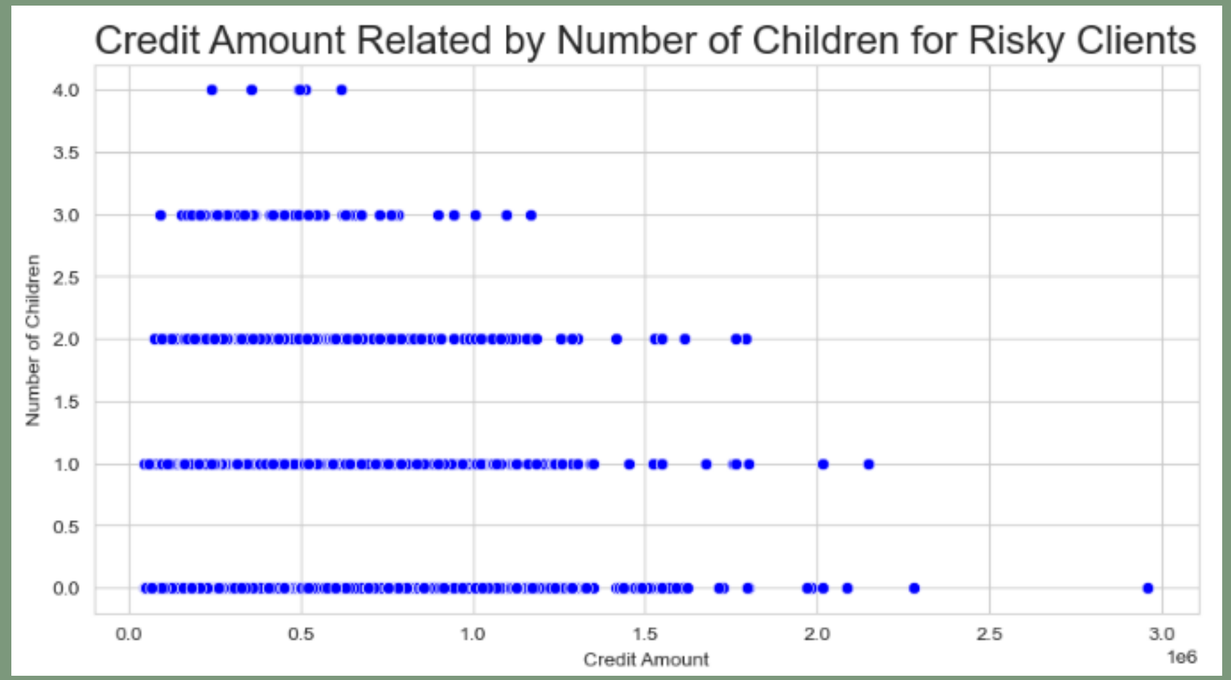


Credit Amount to Number of Children

The credit amount of clients is more likely to be higher the more children they have.

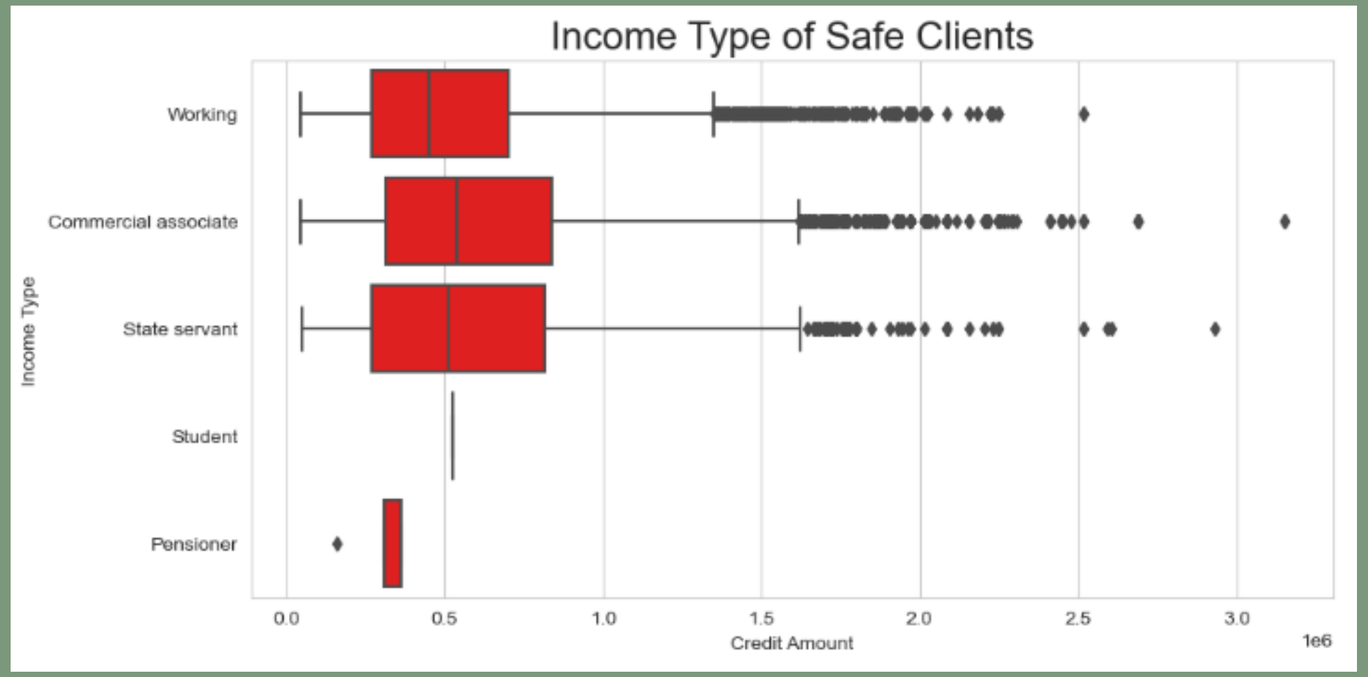
*Clients were categorized to two groups
TARGET 1 (have difficulty repaying the loan) and TARGET 0 (all others)
*They have been renamed for the clarity of the presentation
TARGET 1 = Risky Clients
TARGET 0 = Safe Clients

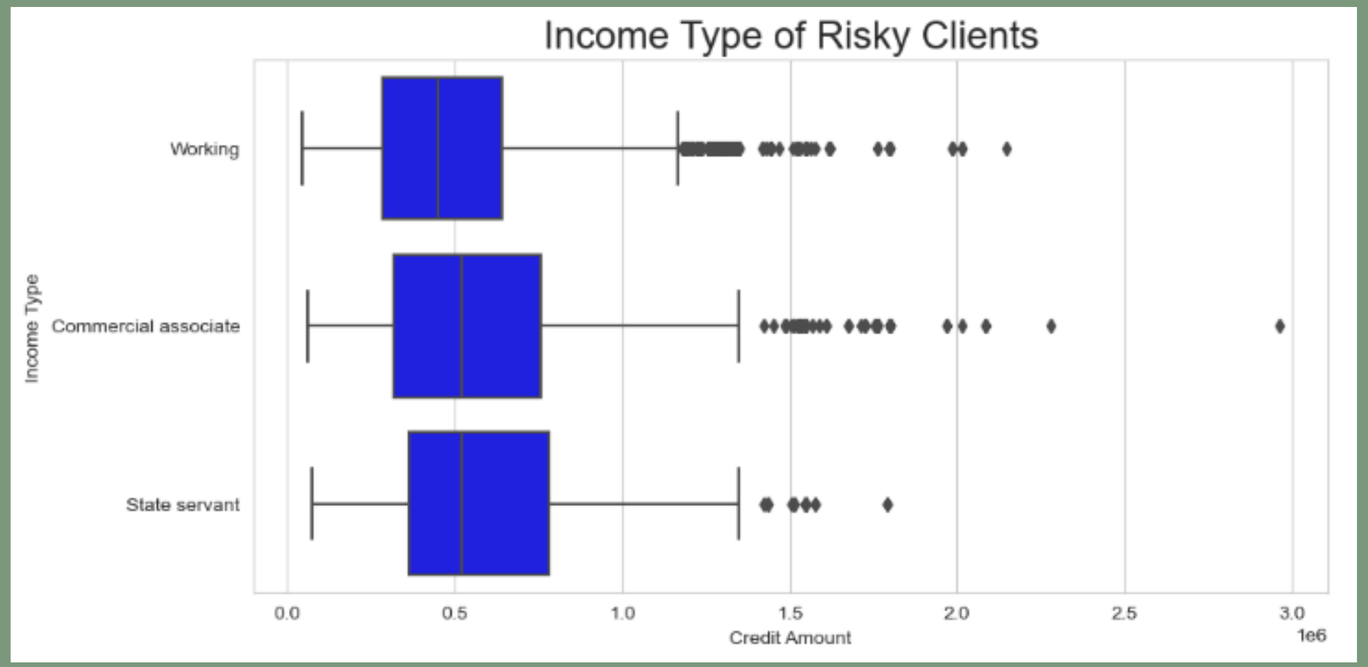




Credit Amount to Income Type

Commercial Associates are more likely to be safe clients.
State servants are more likely to be risky clients





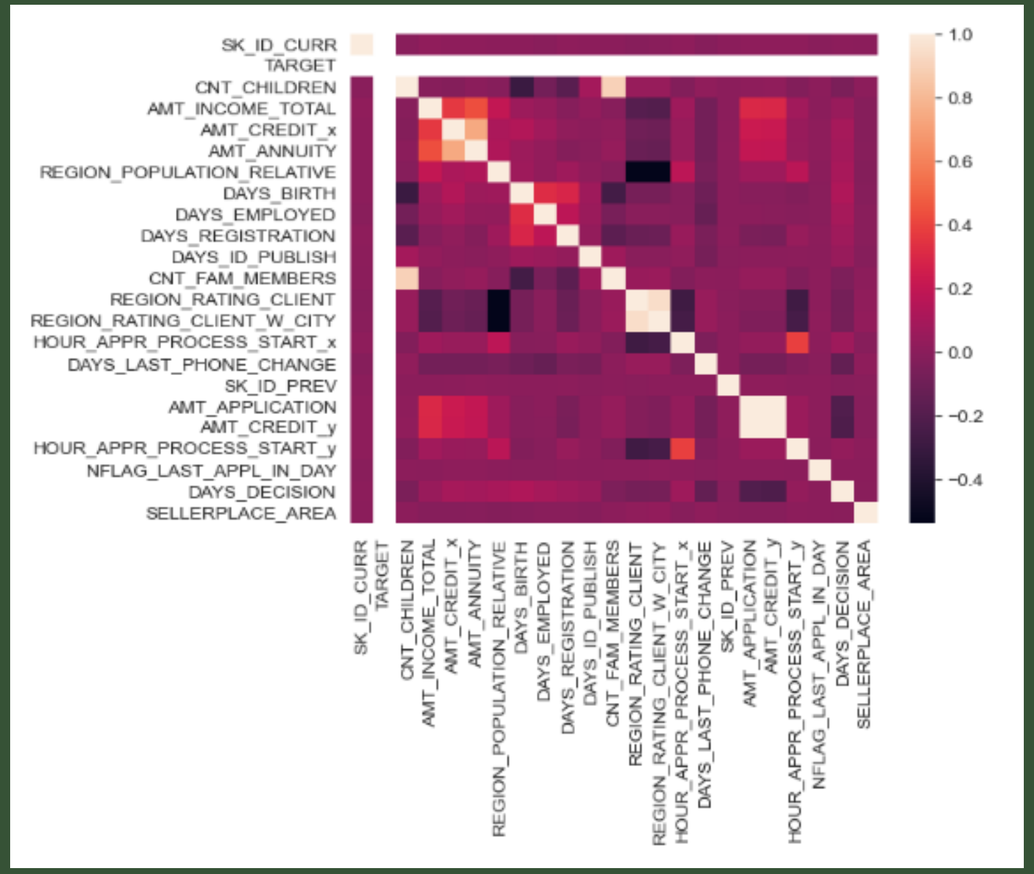
Family Status to Annuity Amount

There is a strong indicator where single clients are more likely to be risky clients.
Married clients were both majority in safe and risky clients.

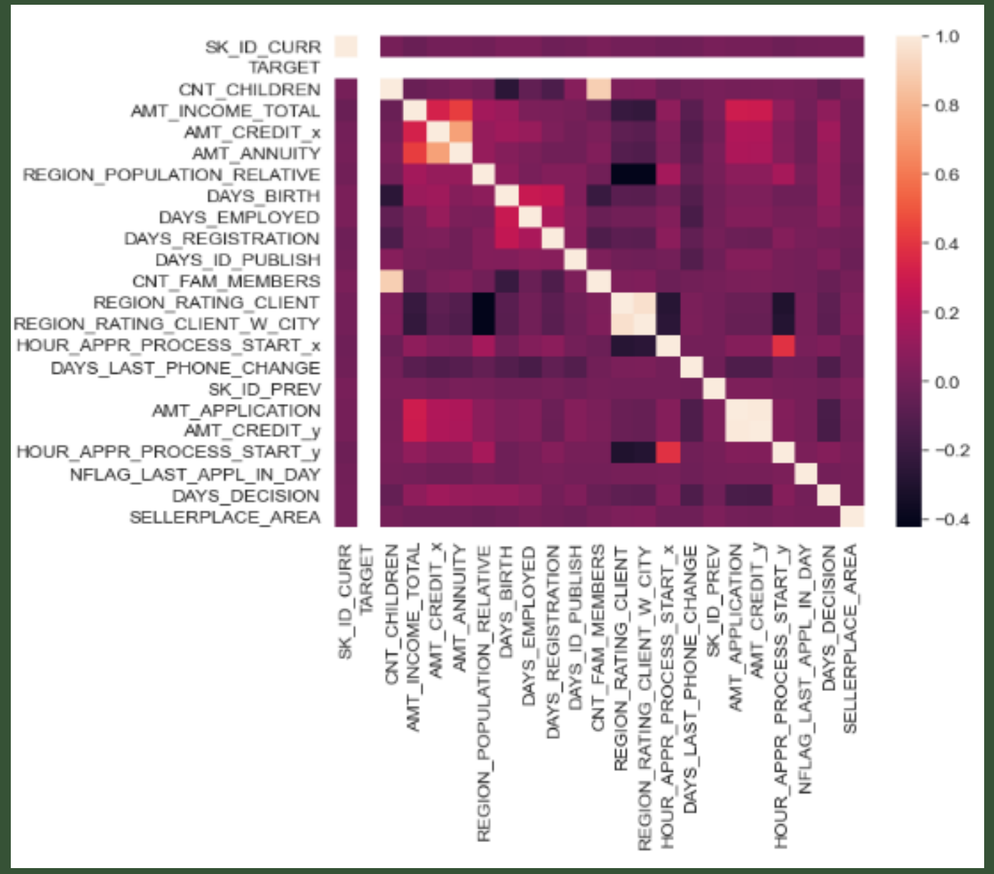
Heatmap Analysis

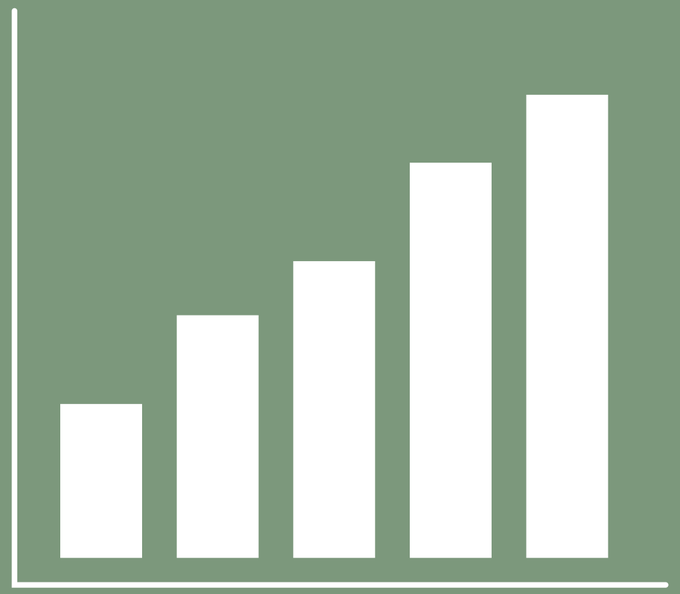
Heatmap for Risky and Safe Clients

Target 0 (Safe)



Target 1 (Risky)

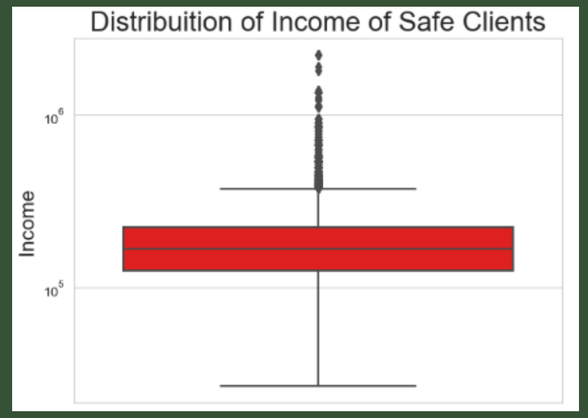


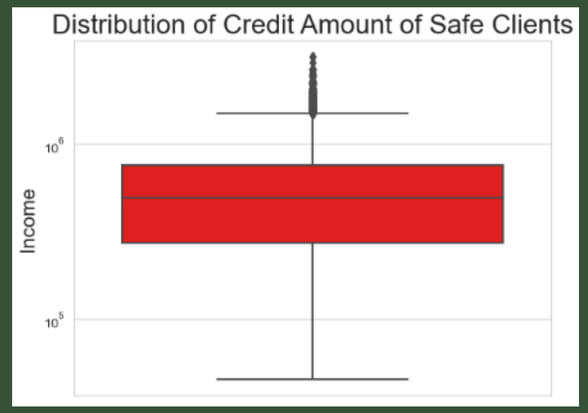


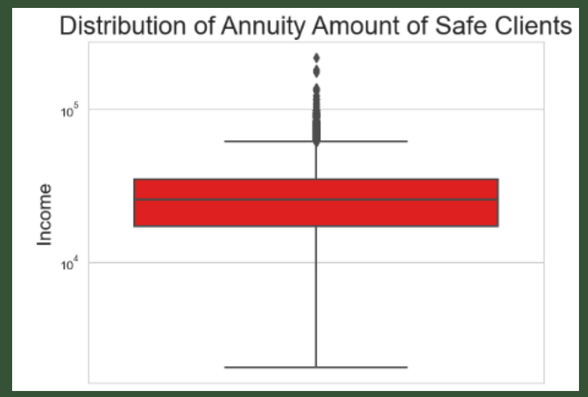
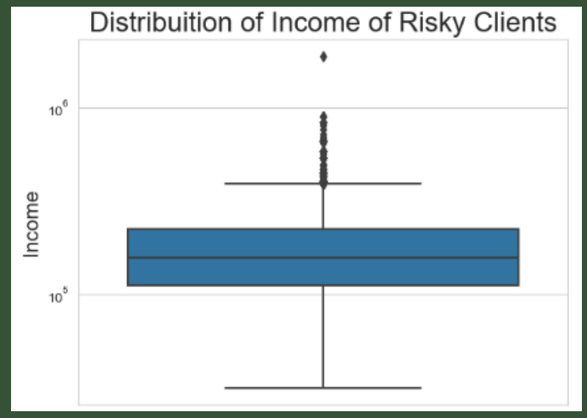
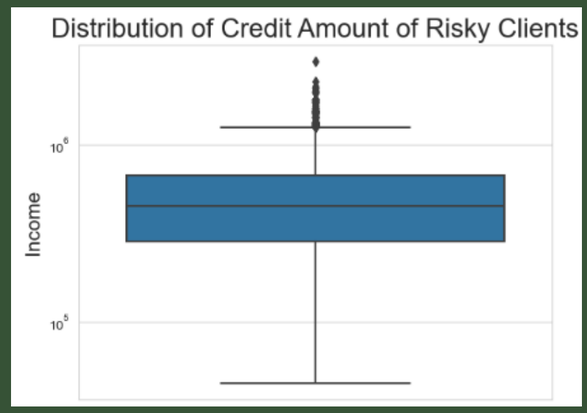
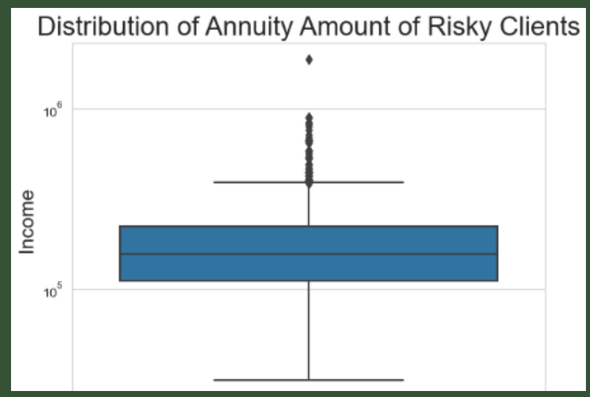
Outliers

Outliers

Safe Clients







Risky Clients