By Sachin Hase

Credit EDA Assignment

Problem Statement

Loan providers struggle to lend to individuals with insufficient or no credit history.

Some consumers exploit this by defaulting on loans.

As an employee of a consumer finance company focused on urban customers, use EDA to analyze data patterns.

This analysis helps identify applicants capable of repaying loans to avoid unnecessary rejections.

Loan approval decisions are based on the applicant’s profile, with two associated risks:

- Not approving a likely repaying applicant results in lost business.
- Approving a likely defaulting applicant leads to financial loss for the company.

Objective:

- The study aims to identify driving factors behind loan defaults (driver variables).
- These variables serve as strong indicators of default risk.
- Understanding these factors aids in portfolio management and risk assessment.

Methodology Used

Data Import

Data Cleaning (delete of columns based on missing value and irrelevant)

Impute missing value for categorical and numerical columns.

Outlier Treatment – Capping and Flooring, Binning

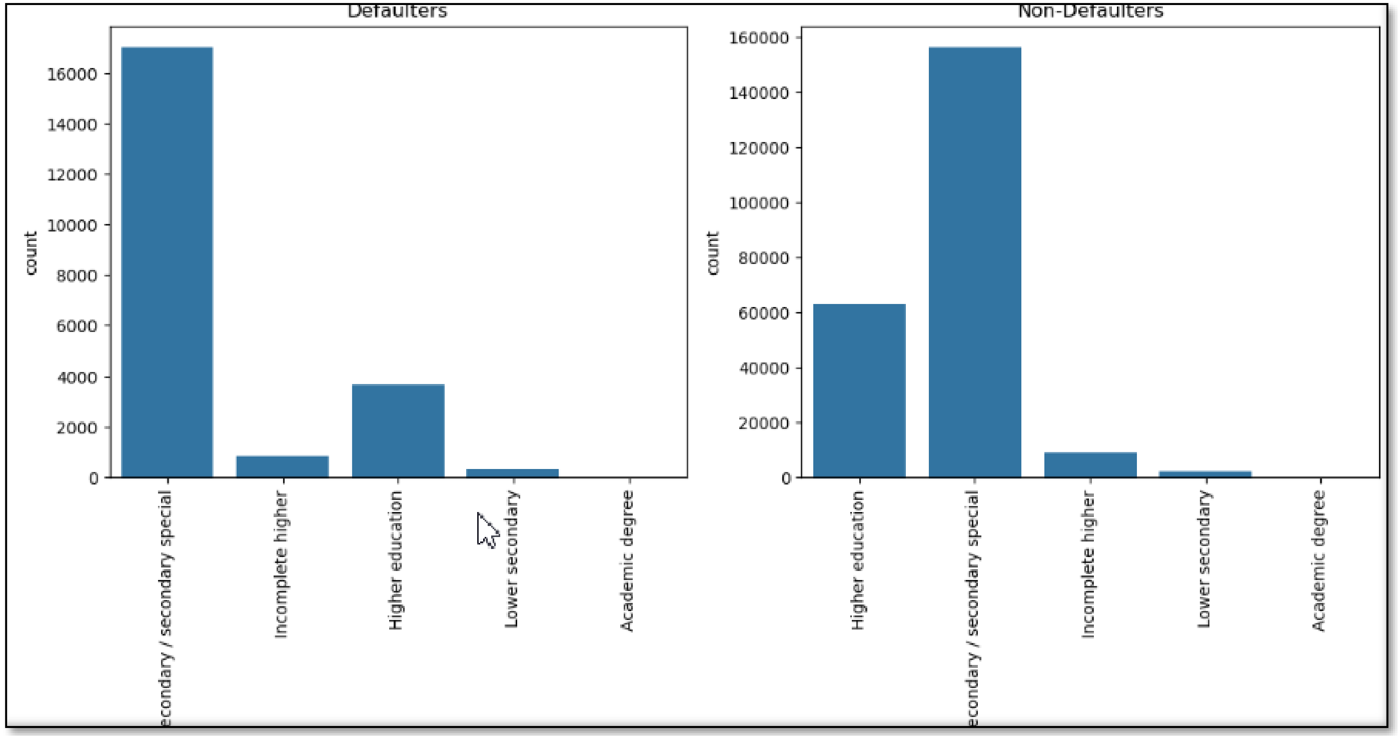
Data Imbalance check

Merging two Datasets

EDA Analysis (Univariate, Bivariate & Multivariate)

Current Application

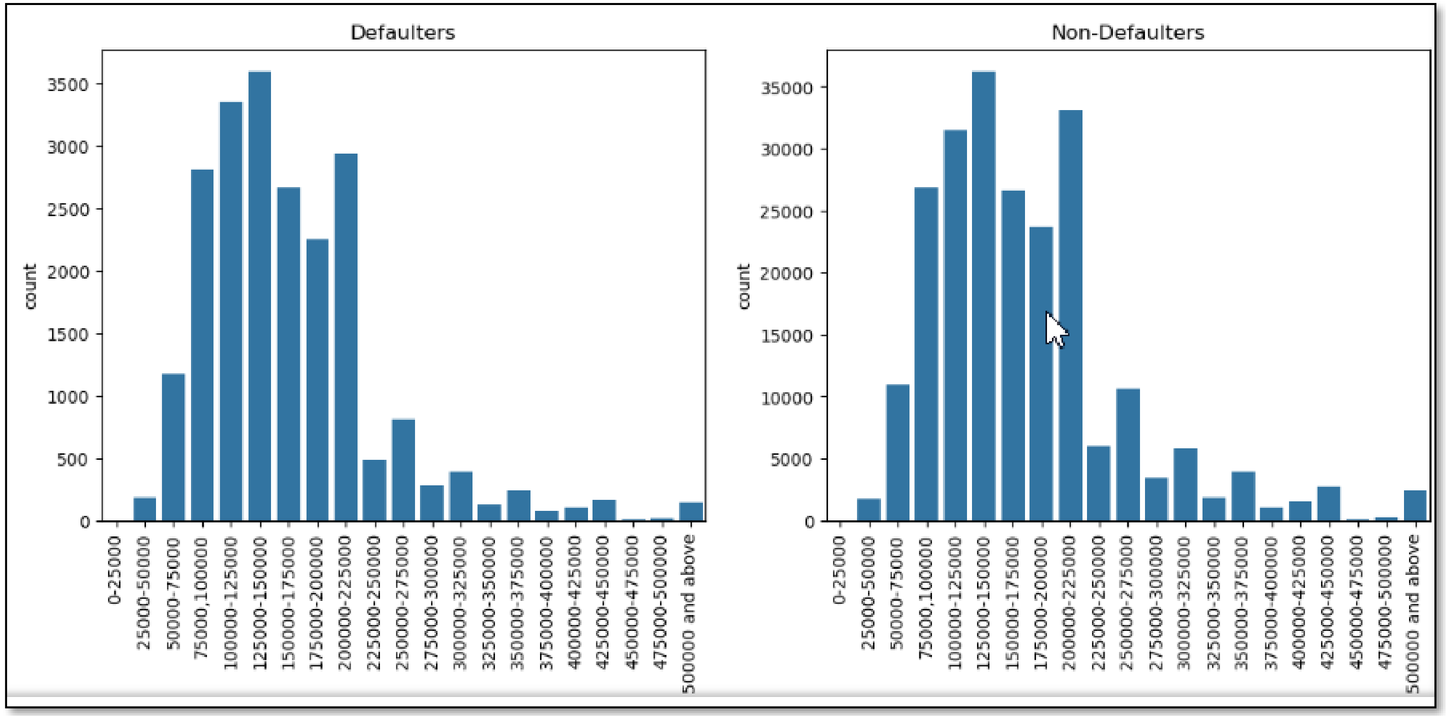
Education Type and Defaulters count



Observations:
1) Secondary/secondary special education type people are more defaulters

Current Application

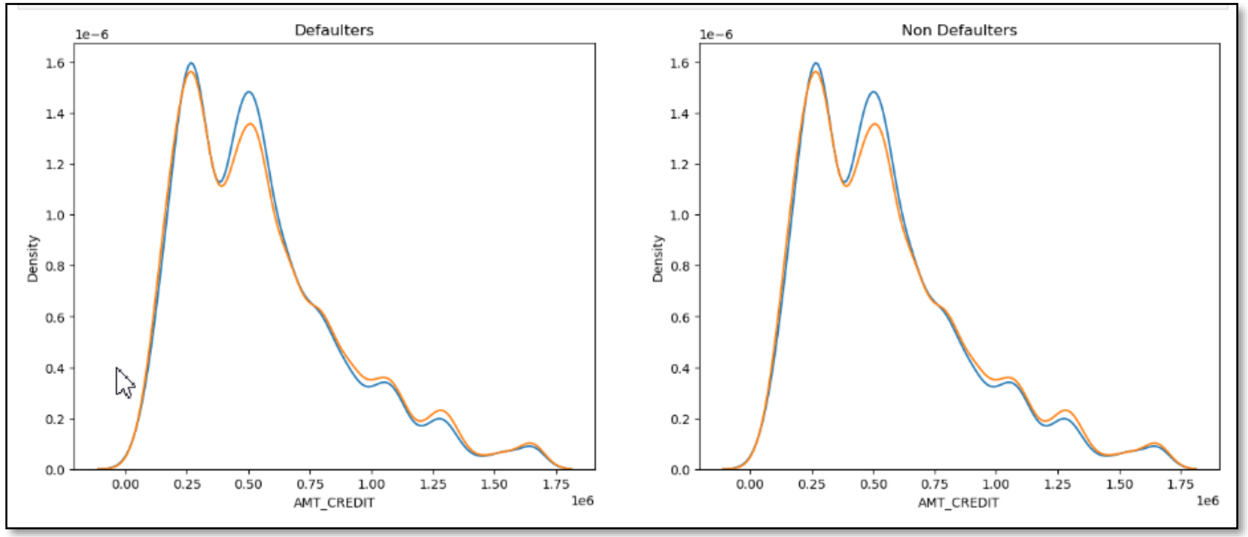
Salary Range and Defaulters count



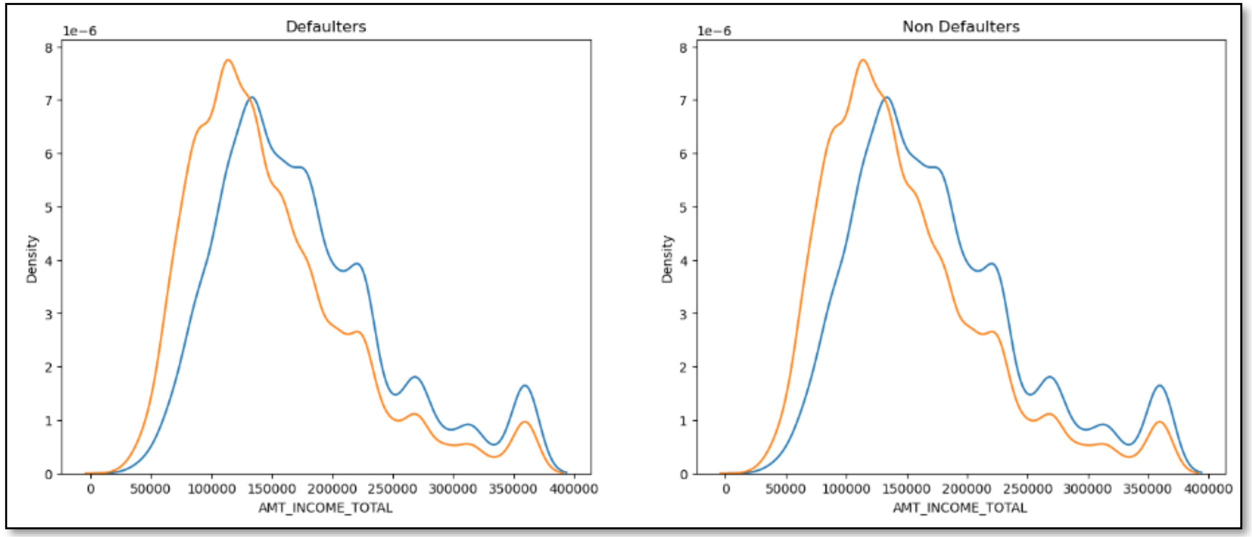
1) People who has salary range(75K to 225K) are seems more defaulter

Current Application

Credit Amount Distribution Vs Defaulters Type



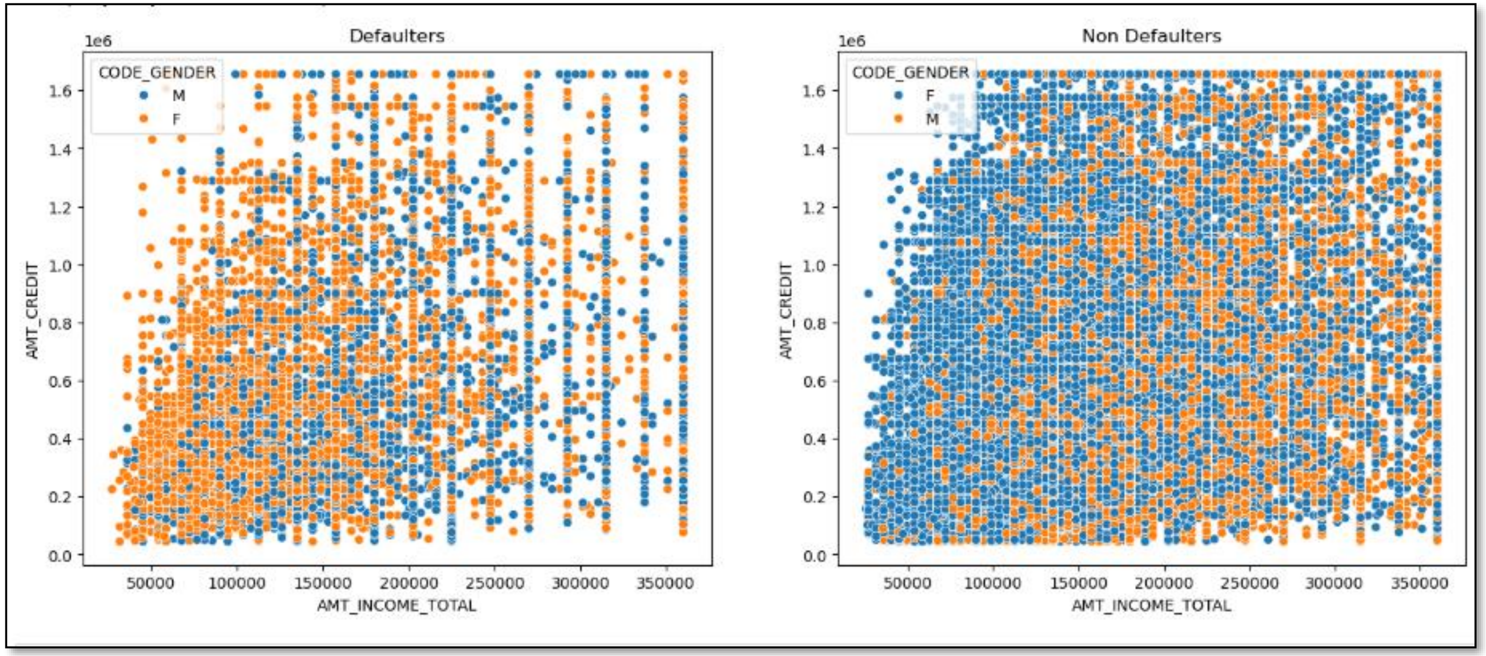
Income Amount Distribution Vs Defaulters Type



1) We can notice that the lesser the credit amount of the loan, the more chances of being defaulter. The spike is till 500000.
2) If the credit amount is less, there is lesser chance of being defaulted. And gradually the chance is being decreased with the loan credit amount.

Current Application

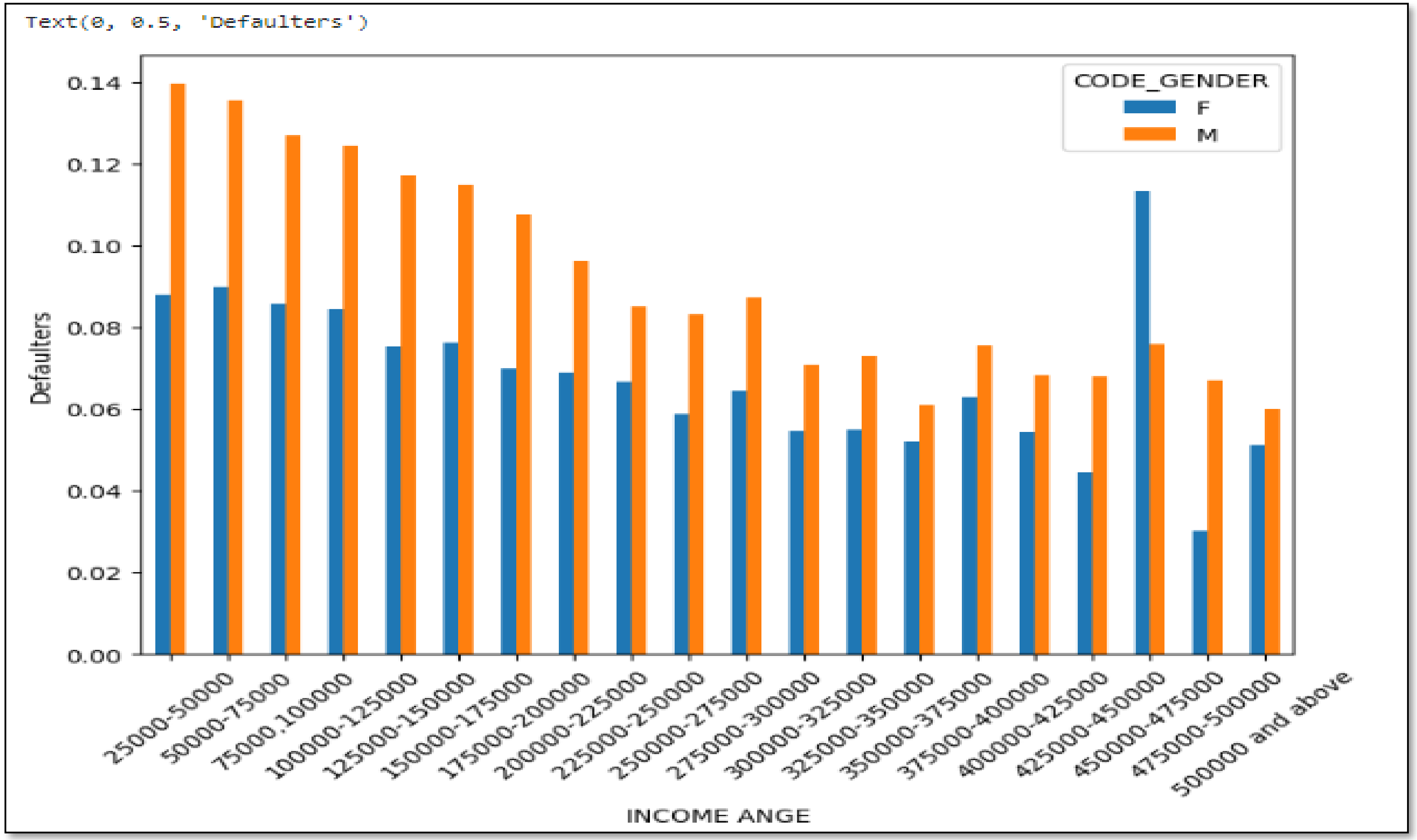
Income Amount and Credit Amount relation



Defaulters –
1) More defaulters are at lower range of Amount Income Total for less loan amount

Current Application

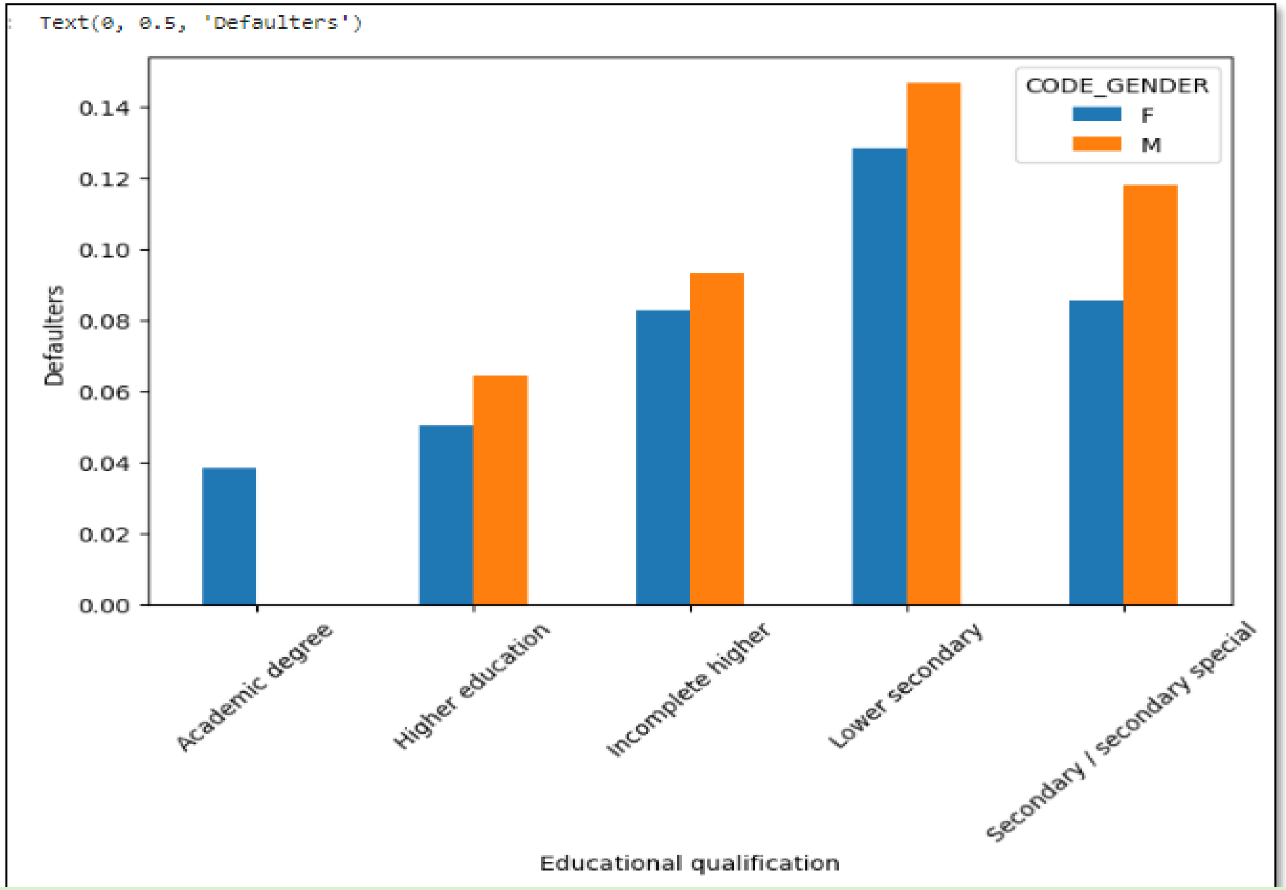
Income Range with Gender vs Defaulters



Observation: Males are more likely defaulted than Females across all income range.

Current Application

Education Type with Gender Vs Defaulters



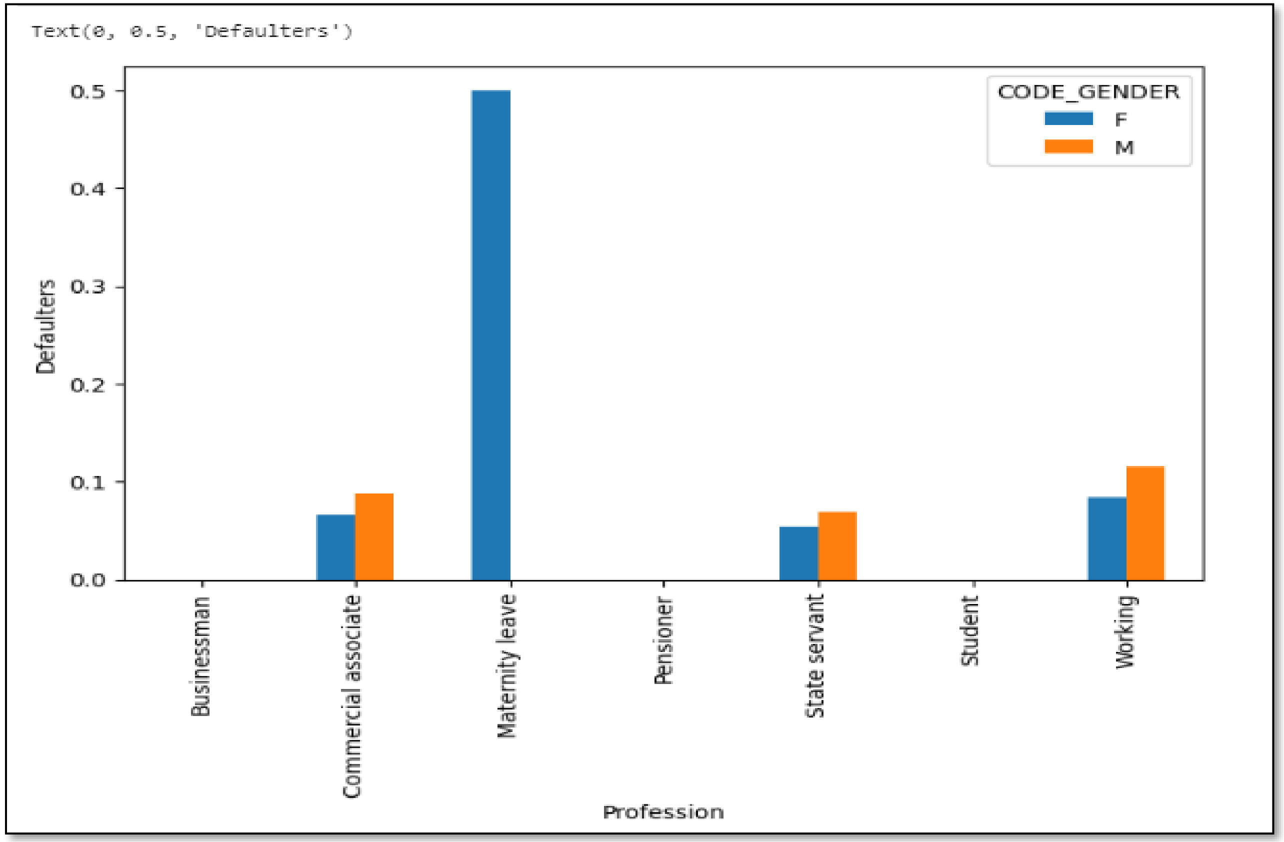
Observations:

1)Lower secondary educated clients are more defaulted followed by Secondary and Incomplete higher educated clients.

2)The Higher educated people are less defaulted.

Current Application

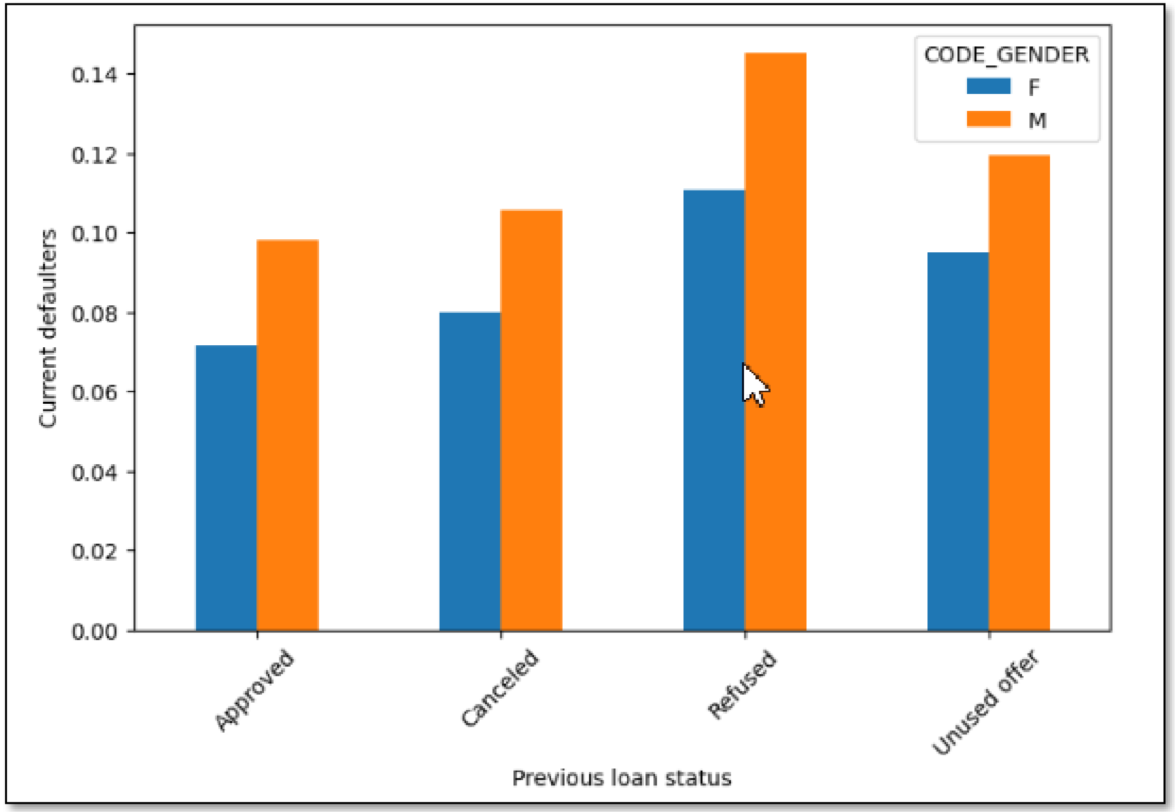
Profession Vs Defaulters



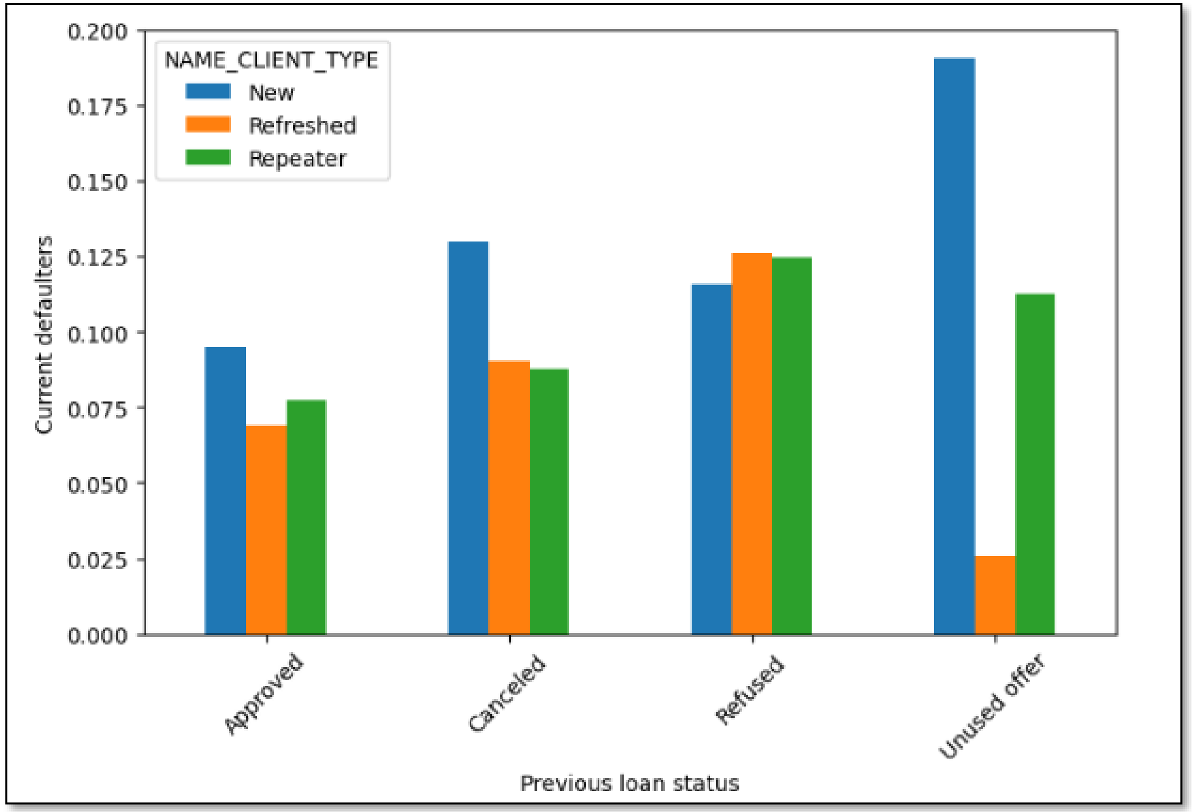
Observation:
1) Clients with maternity leave are expected to be defaulted more.
2) The default rate is lesser in all other professions.

Previous Loan Status Relation with Current Application (Merged Dataset)

Loan Defaulters Across Status



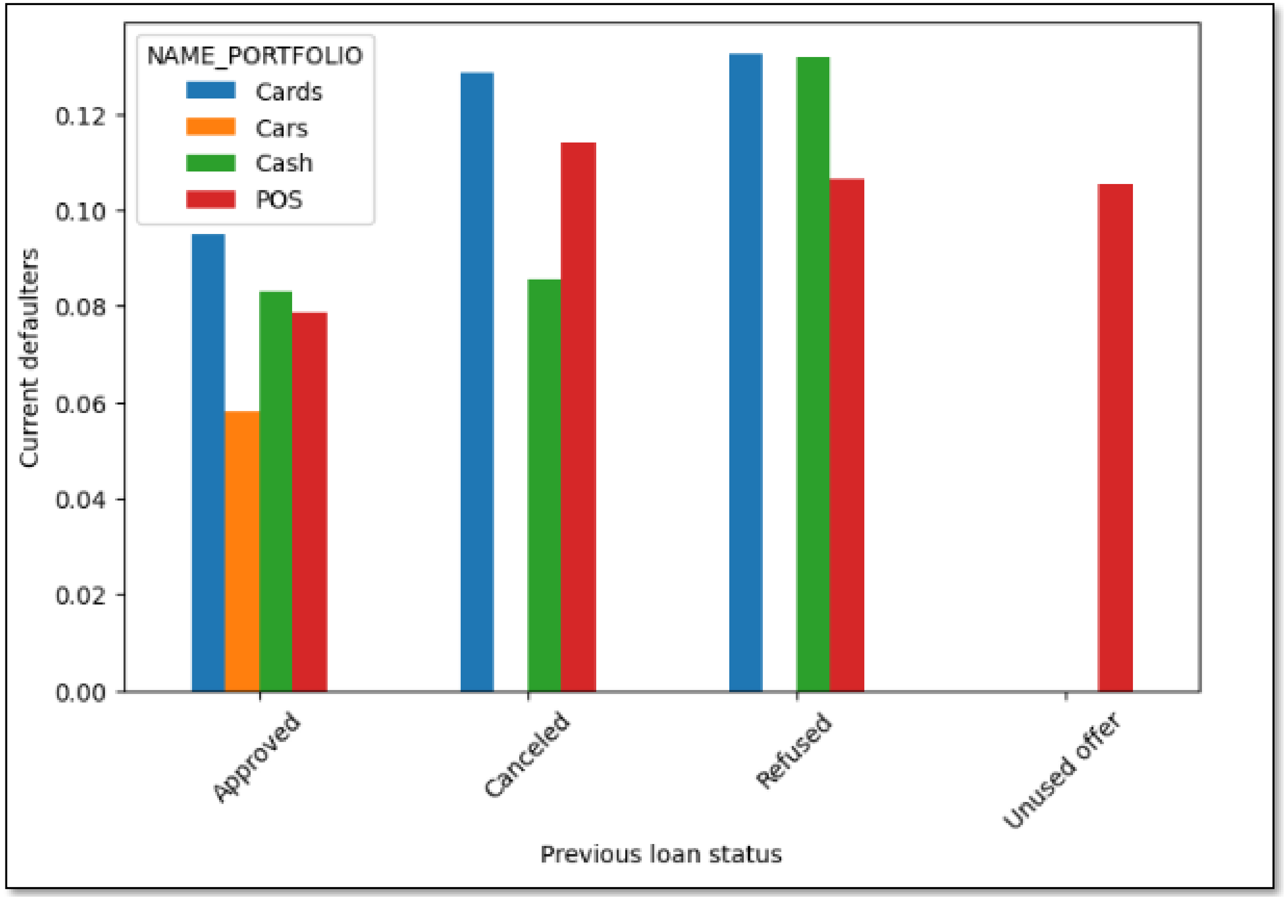
Loan Status and Client Type



previously Refused client is more defaulted than previously Approved clients.

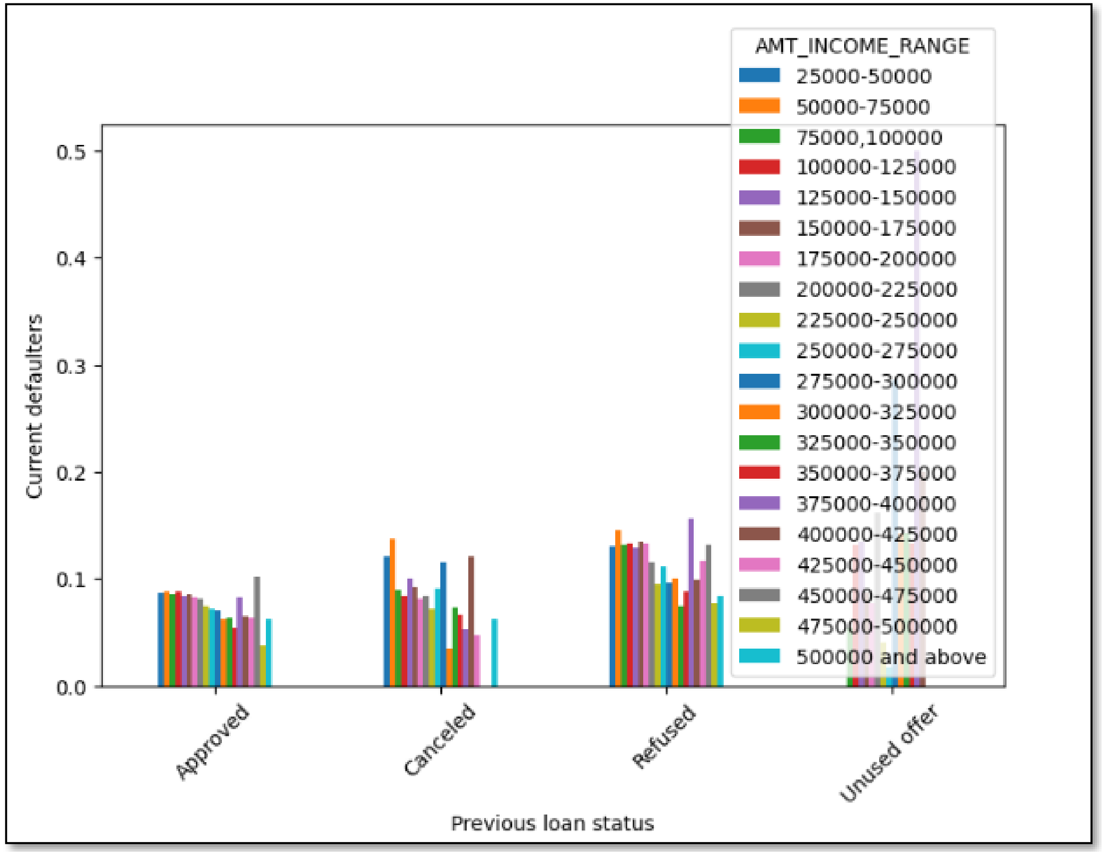
Defaulters are more for previously Unused offers loan status clients, who were New.

Previous Loan Status & Portfolio



Cards and POS are mostly defaulted
Previously refused clients for cash are also have defaulter rate high.

Previous Loan Status and Income Range



Previously Unused offer the Medium income group was more defaulted and Low-income group is the least.

Conclusion:

- The columns (Driving factors) which states information about **previous application Status, Education Type, External Source score, Female Status and gender Status** are important to predict defaulters.
- **High Risk groups are**:
  - Lower Secondary, Secondary/secondary special education type.
  - salary range(75K to 225K)
  - lesser credit amount of the loan
  - Male clients
  - Female Clients with maternity Leave
  - Previously refused or unused clients are more tend to default
  - Cards and POS portfolio.