

CREDIT EDA ASSIGNMENT

SHARMILA SHERIN

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.

When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile.

Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

DATA

The data contains the information about the loan application at the time of applying for the loan and previous application .

1.The client with payment difficulties: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,

All other cases: All other cases when the payment is paid on time.

2. When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

Approved, Cancelled, Refused, Unused offer.

OVERALL APPROACH

Data Cleaning

- Columns having more than 40% missing values were removed.
- Then columns that seemed irrelevant for analysis were identified and dropped.

Missing Values Treatment

- Category type columns were imputed with most frequent values.
- Numerical columns were imputed with median as the outliers were detected.

Outliers

- Outliers were detected , in some of the columns, outliers could be present due to data entry errors and would require further analysis.
- Suggesting, to treat with the upper or lower limit values in numerical column types

OVERALL APPROACH

Data Standardization

- Ensuring all observations under one variable are expressed in a common and consistent unit.
- Some of the observational columns like flag type with 0 and 1 values were converted to categorical type to be represented as 'yes' and 'no' for better visualization.
- The columns which are negative were changed to absolute.

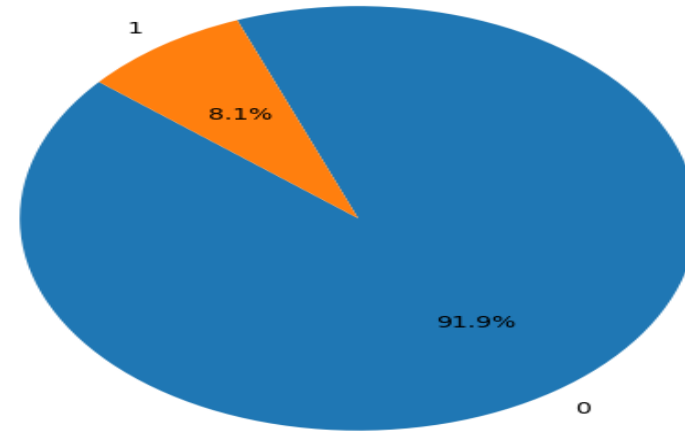
Binning

- Created bins/range for certain columns like Age, Income Total, Family status for better visualization and reaching meaningful insights.

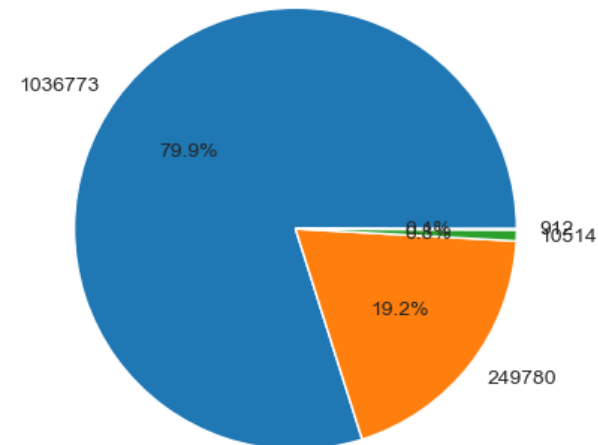
INSIGHTS

- Target data is not balanced, only 8% clients are shown who defaults and 91.9% doesn't default .
- Data imbalance ratio of application data is 11.38%
- Almost 62% applications are approved in previous application data and 18% and 17% are Cancelled and refused respectively.

Distribution of TARGET Variable

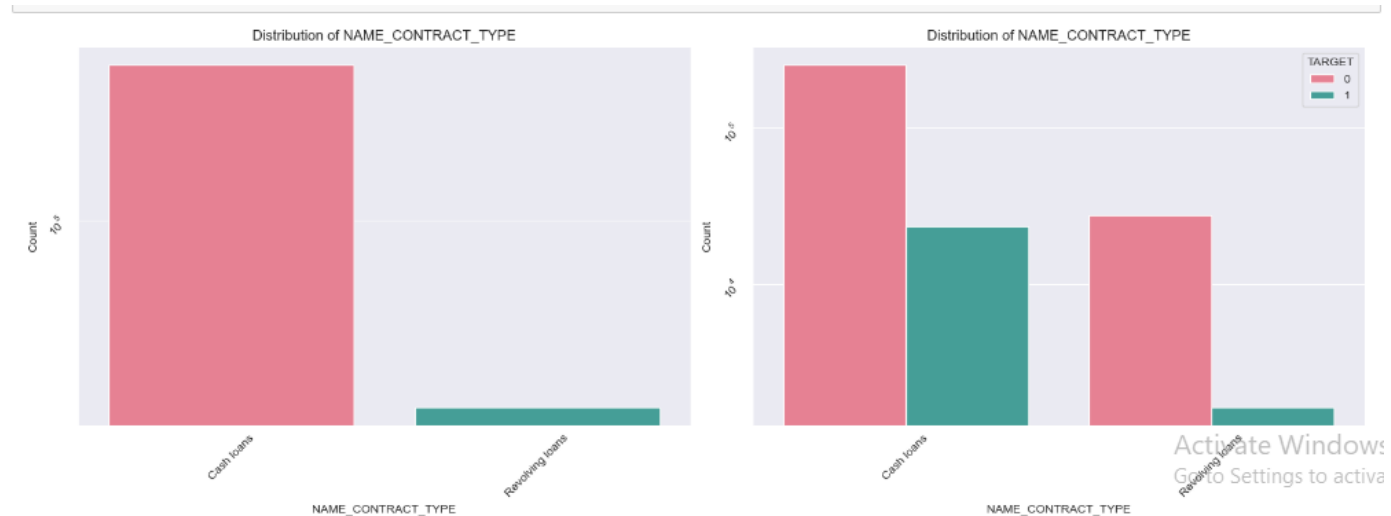


Proportion of Previous Application Status

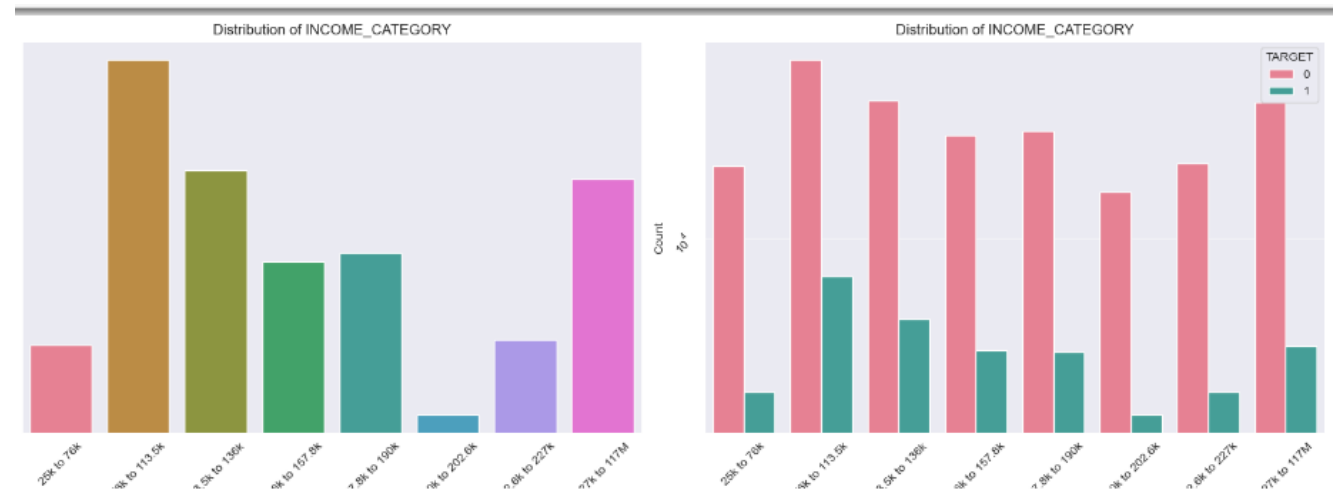


INSIGHTS

- Cash loan Contract type clients are doing defaults compared to Revolving loan type

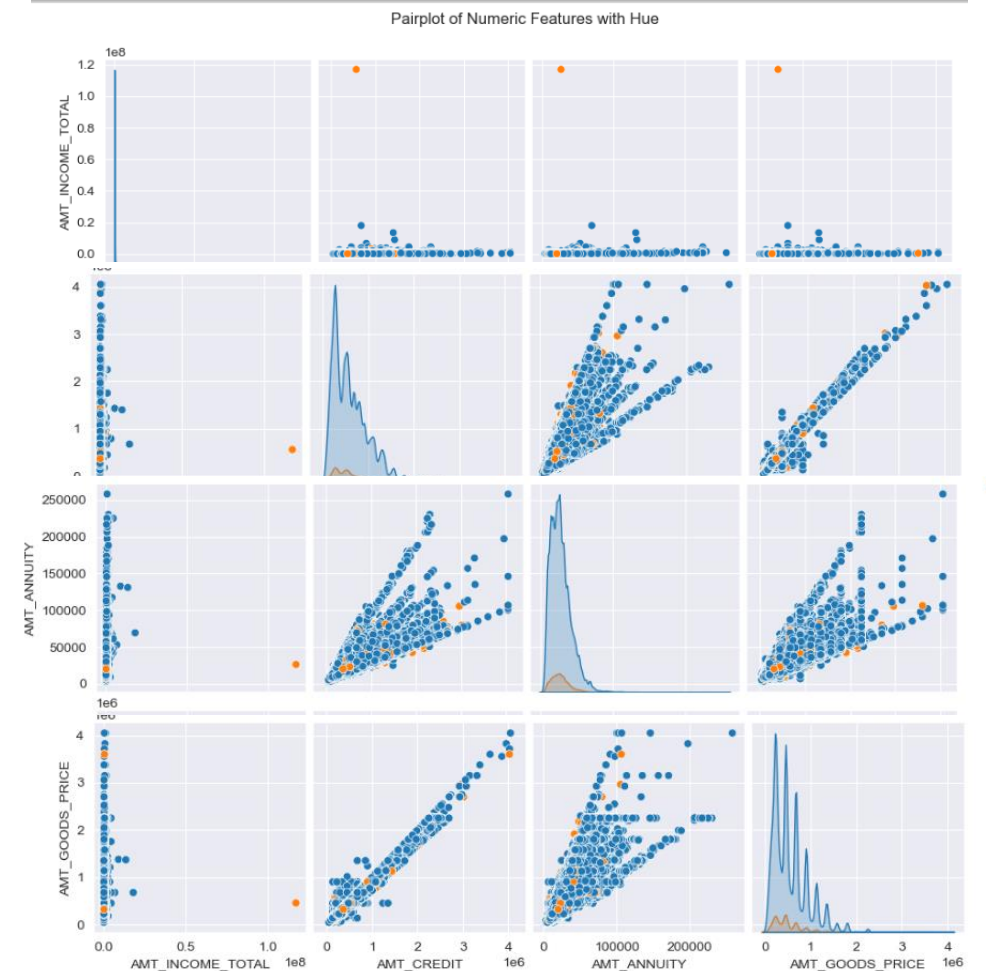


- Majority of people are in 76k to 113.5k income range
- most of the defaulters also lies in this region



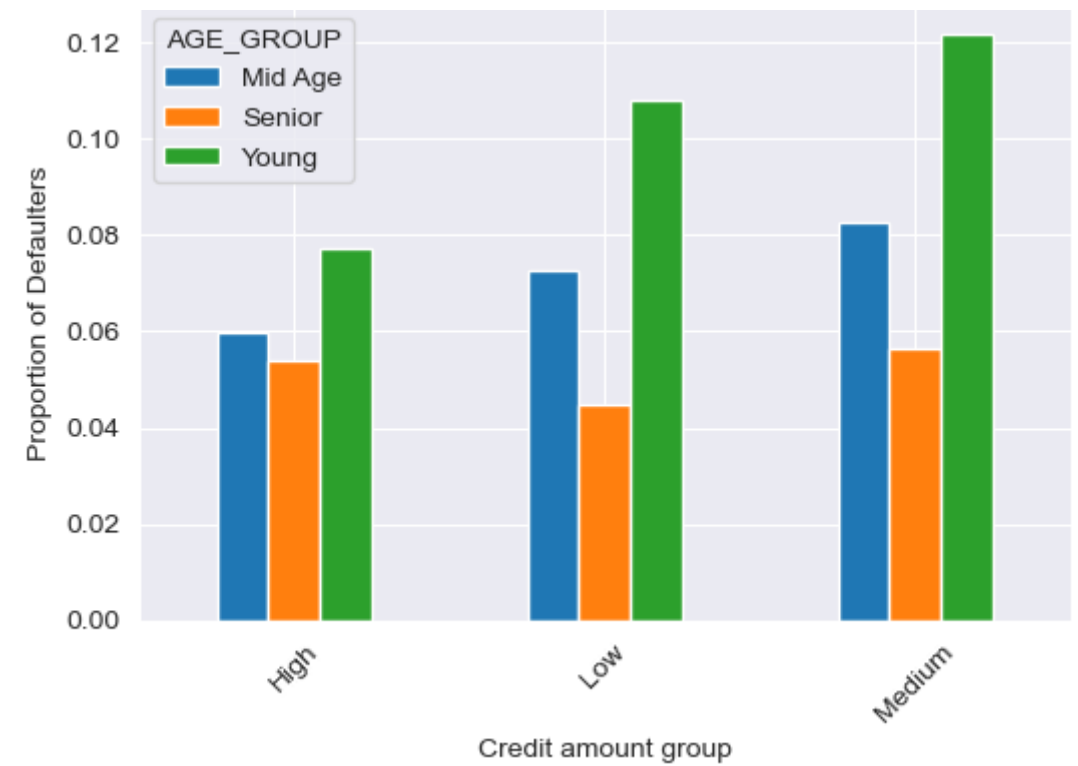
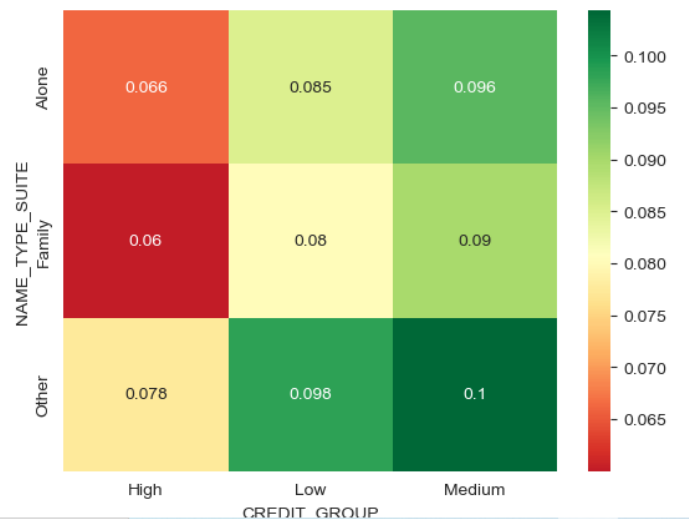
INSIGHTS

- When amount annuity is greater than 15k and amount goods price is greater than 20 lakhs there appear to be a lower chance of defaulters
- This suggests that clients with higher annuity and goods price values may be more reliable in terms of loan repayment.
- Correlation between loan amount and goods price loan amount are highly correlated. We can see in the scatterplot that most of the data points form a line
- There are less defaulters whose loan amount exceeds 2 millions. So whose loan amount is high is less likely default



INSIGHTS

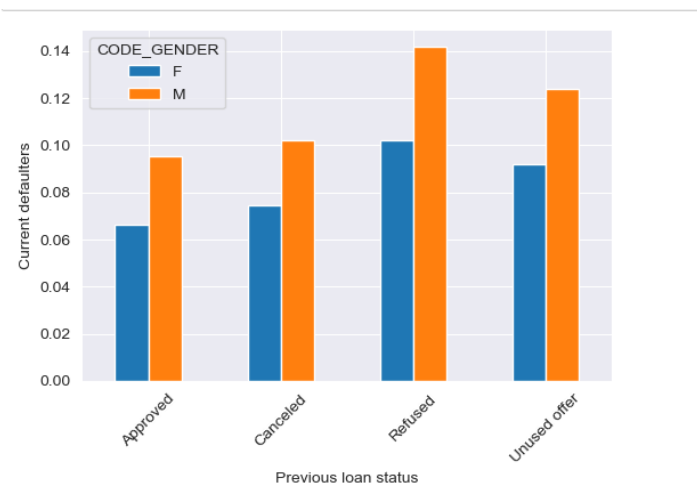
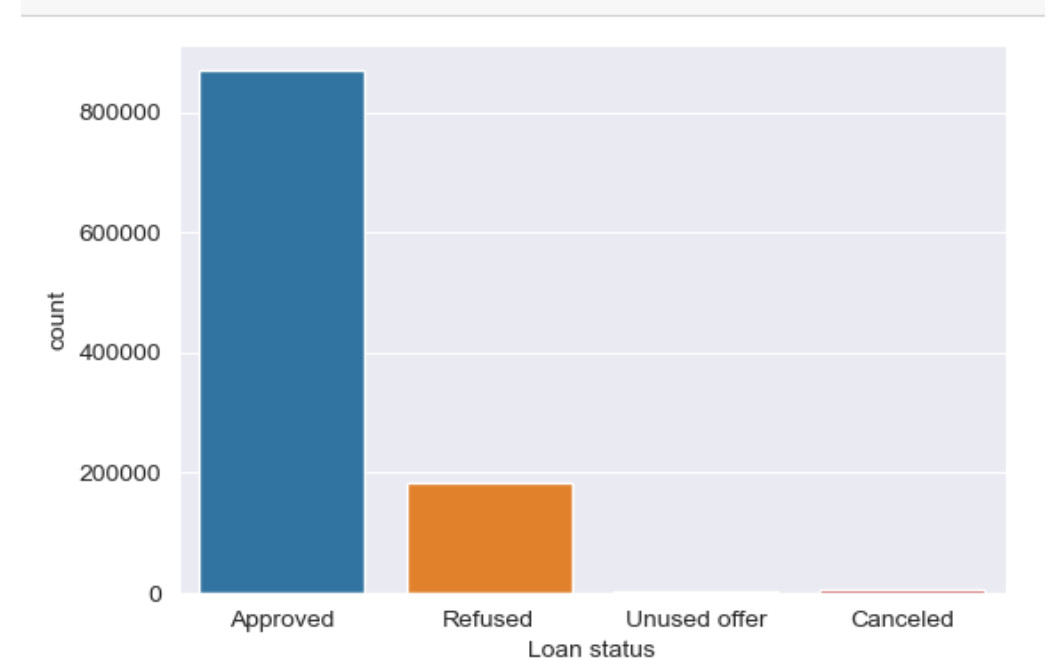
- Young clients with medium and low credit amount group are highly defaulted.
- Senior citizens across all credit amount groups are less likely defaulted.



- people who come alone and other with medium and low credit are tend to default.

INSIGHTS

- There are huge number of Approved loan than Refused.
- Hardly, there are any Canceled or Unused offer loan



- We see that previously Refused client is more defaulted than previously Approved clients. Also, in all the cases the Males are more defaulted than Females.

SUMMARY

Factors indicating possibility of loan default:-

- **Gender Dynamics:** Male applicants are more prone to default
- **Marital Status Matters:** Individuals in civil marriages/single status has higher default rate.
- **Education level :** Less educated people are more likely to default their loan.
- **Loan history:** Applicant whose previous loans are refused/cancelled are more likely to default their loan.
- **Occupational Hazard:** Unemployed, unskilled staff like labors, and people with less stable jobs are more likely to default.
- **Youthful Tendencies:** Younger applicants aged between 20–40 face a comparatively higher probability of default.
- **Income Anomalies:** Clients on maternity leave or experiencing unemployment tend to have a higher likelihood of default.
- **Family Size Influence:** Clients with larger families or more children tend to have higher default rates.
- **Work Experience Variable:** Clients with less than five years of employment history exhibit an elevated default rate.

RECOMMENDATIONS

- Inspecting the reasons of loan cancellation and record loan cancellation reasons could give a chance to renegotiate terms with client.
- Previously rejected clients have successful repayments. Briefly analyzing and reconsider the rejected applicants could mitigate business losses and expand lending opportunities.