



Loan Default Risk Analysis

Analysis by Pritam Chowdhury

Data cleaning and pre-process

Numeric Missing Value:

- The data Provided has substantial missing values in numerical columns which can not be discarded.
- Missing values are imputed with mean values calculated for each group defined by another relevant column like loan type

Categorical Missing Value:

- The data Provided has substantial missing values in categorical columns which can not be discarded.
- Missing value in categorical columns are imputed with mode calculated for each group defined by another relevant column like loan type.

Skewed Data

- In Numerical columns there were multiple outliers
- Outliers are detected by IQR method
- Outliers are handled by clipping the data from $Q1 - IQR \times 1.5$ to $Q3 + IQR \times 1.5$

Analysis Process

Check the distribution

Once data is cleaned and outliers are handled the distribution for the column is checked for both the default loan and non default loan

Default percentage for a group

For categorical columns, default percentage of each group is calculated

Hypothesis test

With the help of hypothesis testing framework, insights from the distribution and charts are validated statistically with 99 percent confidence.

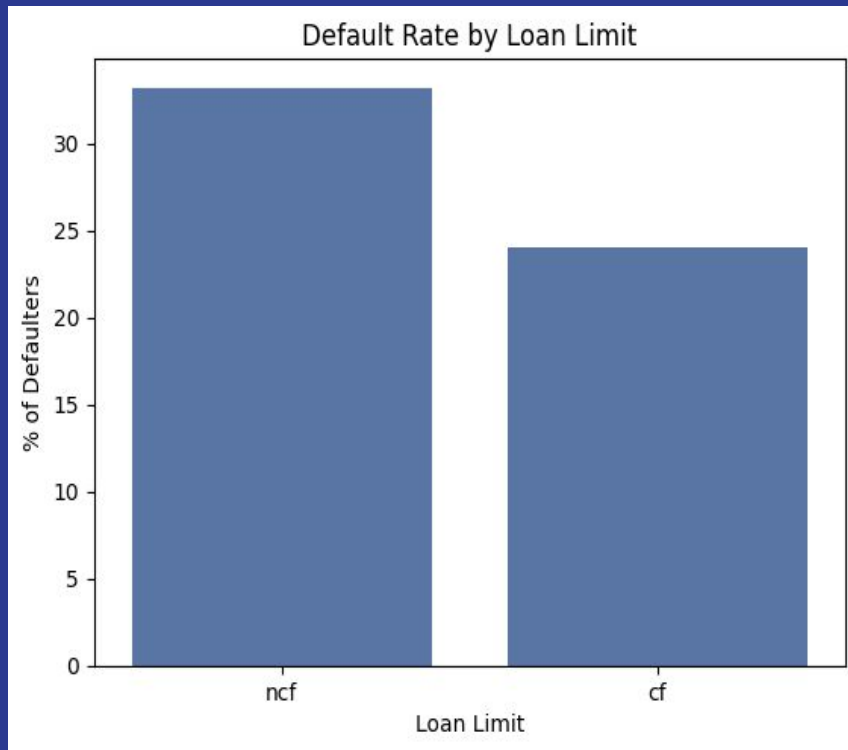
Key Insights

Loan Limit

Loans with no fixed limit has
higher default rate.

Default rate of loans with no fixed limit 33%

Default rate of loans with fixed limit 24%



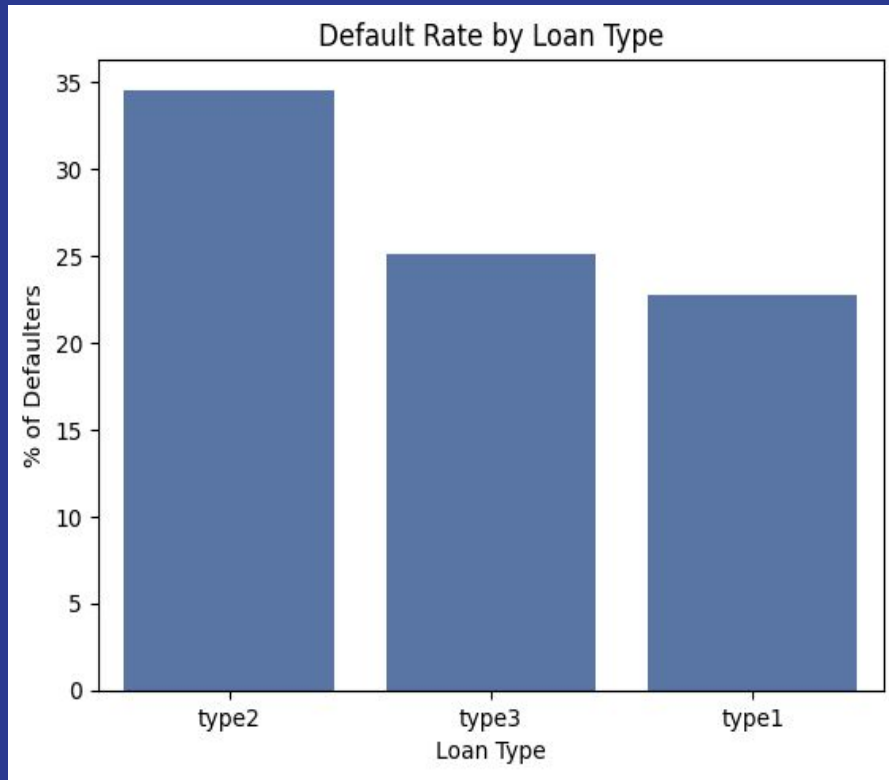
Confidence level ~ 99.99%

Loan Type

Type2 loan has higher default rate whereas Type1 has significantly lower default rate.

Default rate of loans Type 2 35%

Default rate of loans Type 1 23%



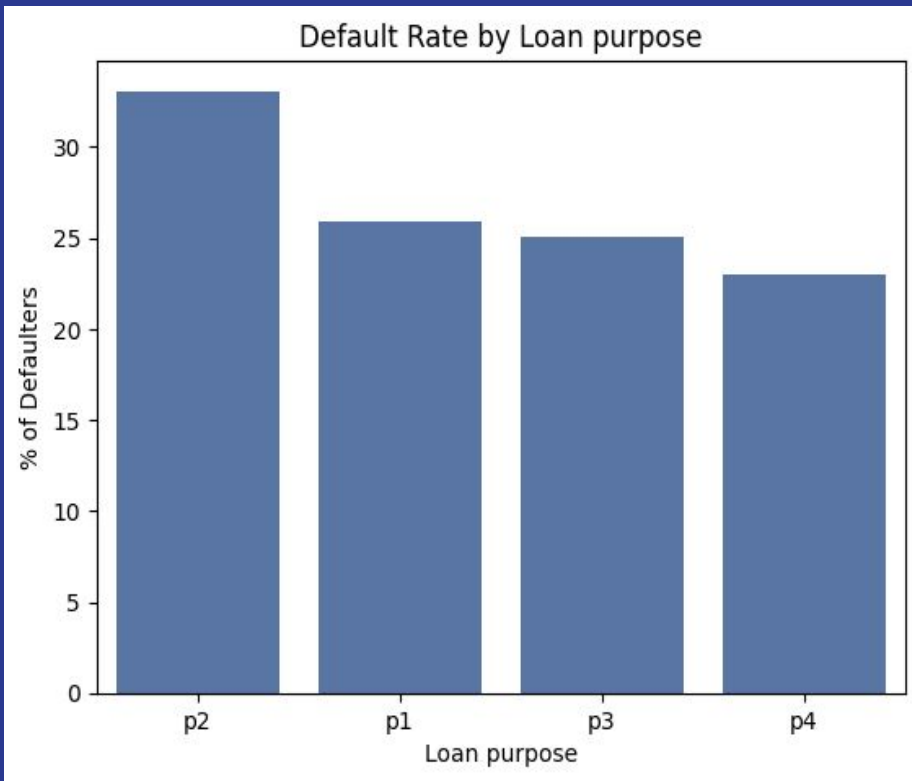
Confidence level ~ 99.99%

Loan Purpose

People who taken loan for p2 has higher default rate and people who has taken loan for p4 are safer comparatively. Business loan has higher default rate.

Default rate of loans purpose p2 33%

Default rate of loans purpose p4 22%

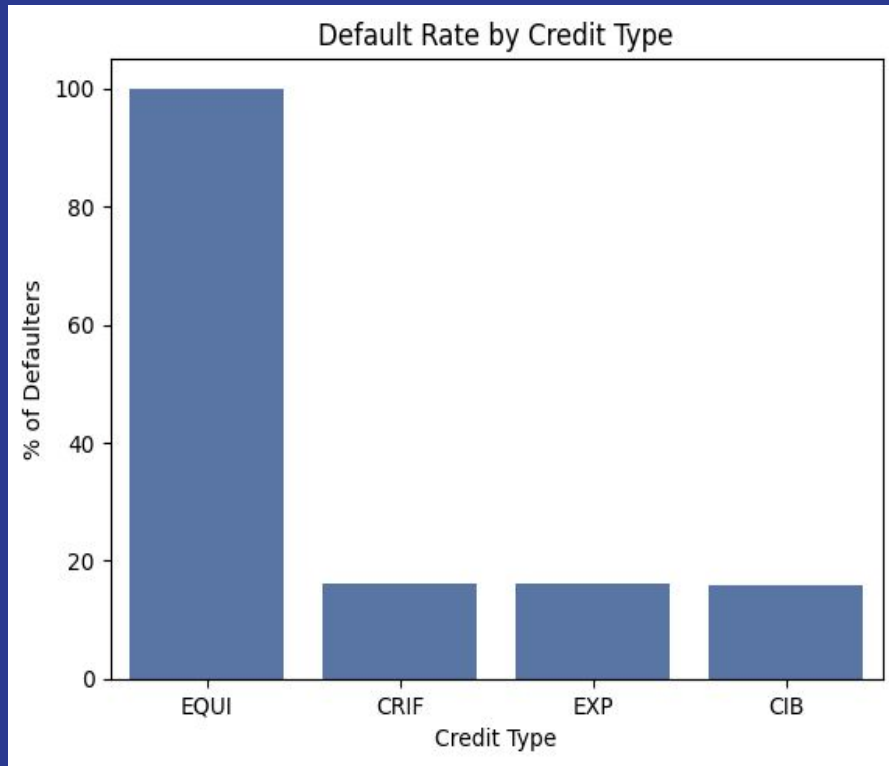


Confidence level ~ 99.99%

Credit Type

For credit type EQUI default rate
is much higher

Default rate of credit type EQUI is 99.99%
Default rate of other credit types are 15% to 16%



Confidence level ~ 99.99%

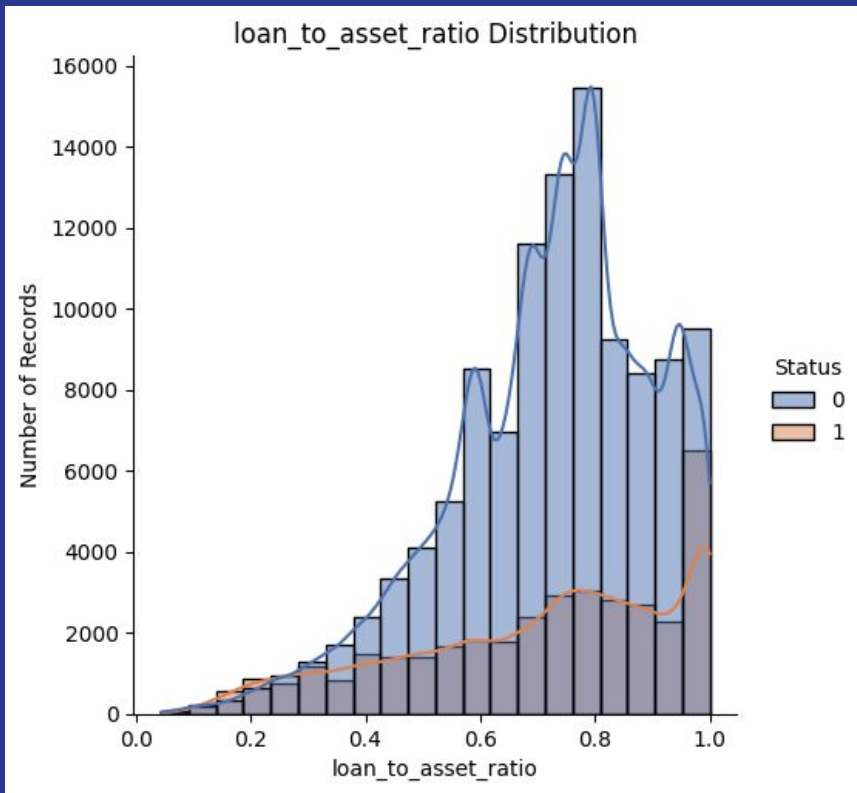
Loan to Asset Ratio

People with high loan to asset ratio tend to less default on loans

Average loan to asset ratio of non defaulters 72%

Average loan to asset ratio of defaulters 70%

Confidence level ~ 99.99%



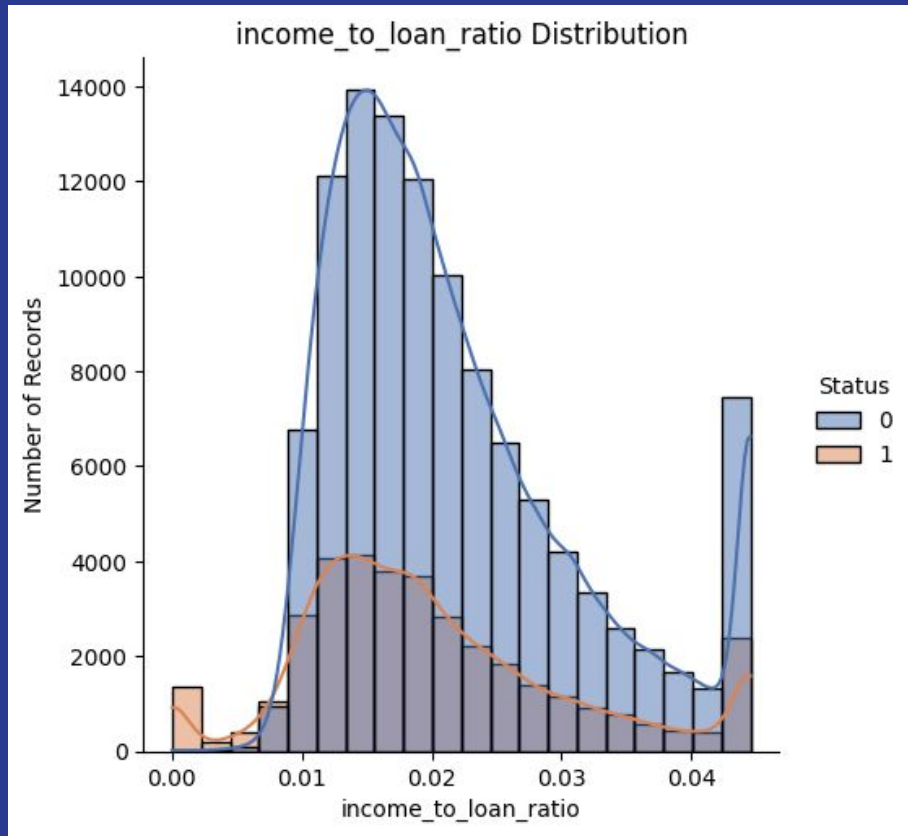
Income to Loan ratio

People with high income to loan ratio tend to less default on loans

Average income to loan ratio of non defaulters 2.2%

Average income to loan ratio of defaulters 2%

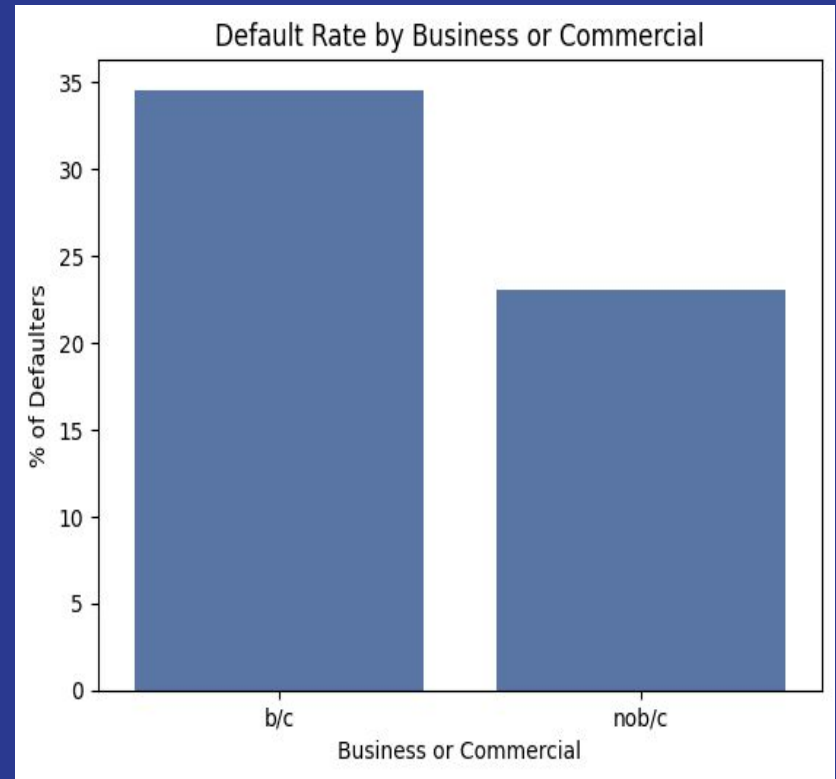
Confidence level ~ 99.99%



Business Loan

Business loan has higher default rate.

Default rate of business or commercial loan is 35%
Default rate of non business or commercial loan is 23%



Confidence level ~ 99.99%

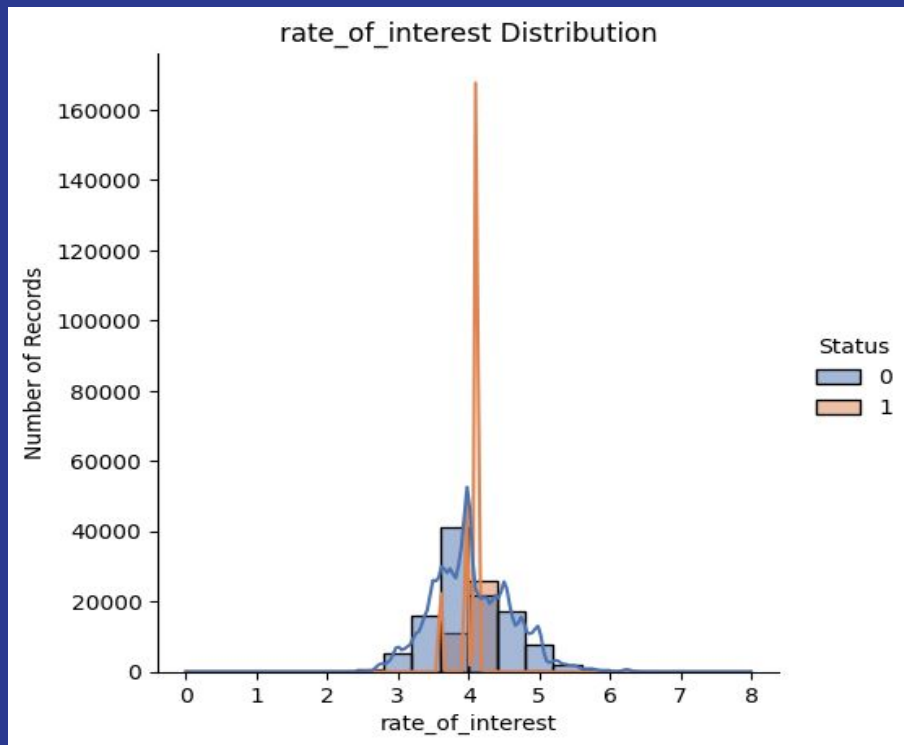
Rate of Interest

Loans with lower interest rate
has higher default rate

Confidence level ~ 99.999%

Average rate of interest of non defaulters 4.045%

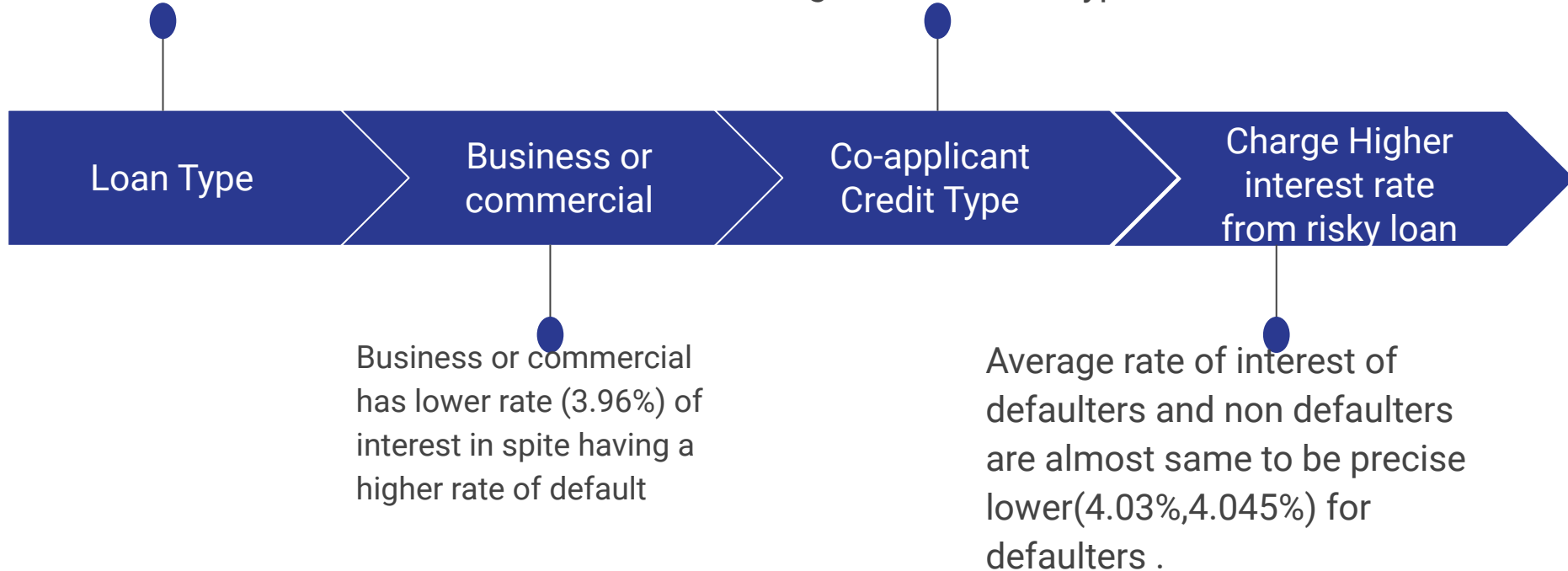
Average rate of interest of defaulters 4.03%



Recommendations

Loan Type 2 has lower rate of interest (3.96%) in spite of having a riskier loan type

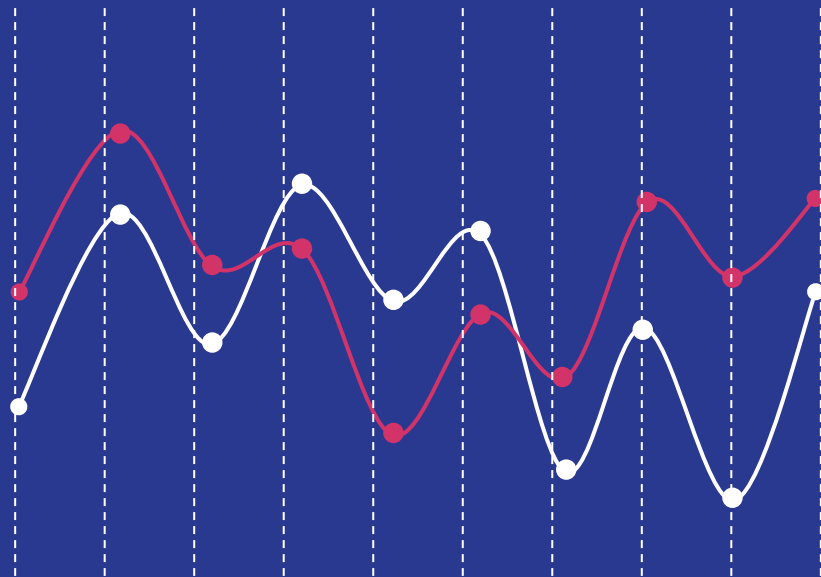
co-applicant_credit_type EXP has lower rate of interest(4.02%) in spite of having a riskier credit type



More Analysis

Notebook link:

<https://colab.research.google.com/drive/1iP2rVDwyq2ERVm9kZrSUsXjanu6VbYFU#scrollTo=eApg-1THyTCs>





Thank you