

Reunion

Insights from given data:

- Given dataset were separated into two different csv files which first needed to be joined using common column which was application id.
- It had 700 records of people having low credit risk and 300 records of people having high credit risk.
- It had many columns with NULL values which were handled as described below
 - Columns which had high number of NULL values were removed since they would affect overall quality of dataset.
 - For columns with categorical values, NULL values were replaced by mode.
 - For columns with continuous values, NULL values were replaced by mean.
- Factors having high probability to be in high-risk zone are:
 - Gender: Female
 - Age group: Student (1822)
 - Purpose: new car
 - Loan amount: Greater than ~15.3 Lakh and also between ~13 Lakh and ~13.8 Lakh
 - Greater the period of loan more chances to be in high risk
 - Account balance: Medium
 - Employment status: management/ self-employed/ highly qualified employee/ officer
 - Housing: For free
 - Marital status: divorced/separated
 - Mean Loan amount taken by Male: 3448040
 - Mean Loan amount taken by Female: 2877774
 - Property: car or other
- To get more insights into data kindly refer here:
<https://colab.research.google.com/drive/1vkqoj-lzdnP8Yyt9MLn6Lgs2ZTXjxi-A?usp=sharing>