HOME LOAN DEFAULT RISK MANAGEMENT

Team ID- PTID-CDS-MAY22-ONL

Ankita Das – anki.das23@gmail.com

Kavita Desai – kdarch14@gmail.com

Prasant Vijay Karn – prasantvkarn@gmail.com

HOME LOAN DEFAULT RISK MANAGEMENT-
INSIGHTS OF DATA

1. **Current Application:-** Static data for all applications. One row represents one loan in our data sample.

2. **Previous Application:-** All previous applications for Home Credit loans of clients who have loans in our sample.

3. **Bureau:-** All client's previous credits provided by other financial institutions that were reported to Credit Bureau (for clients who have a loan in our sample).

4. **Bureau Balance:-** Monthly balances of previous credits in Credit Bureau.

5. **Credit Card Balance:-** Monthly balance snapshots of previous credit cards that the applicant has with Home Credit.

6. **Instalments Payments:-** Repayment history for the previously disbursed credits in Home Credit related to the loans in our sample.

7. **POS Cash:-** Monthly balance snapshots of previous POS (point of sales- **Point of sale** (**POS**) refers to the place where customers execute payments for goods or services. **POS** systems provide companies with sales and marketing data.) and cash loans that the applicant had with Home Credit.

1. BASIC DETAILS:

- No. of defaulters: 24825
- No. of non-defaulters: 282686
- Percentage of defaulters: 8.072881945686495
- It's an imbalance Dataset. (Defaulter : Non-Defaulter = 8 : 92 = 2 : 23)

- Percentage of applicants approved previously but are defaulter in current loan: 7.588655443691958
- Percentage of applicants refused previously but are non-defaulter in current loan: 88.00358612820408
- Most applications are approved. Cancelled and Refused are almost on the same level.

2. CURRENT APPLICATIONS:

- The shape of Current Application: (307511, 122)
- Number of duplicated files are: 0
-

2.1. CATEGORICAL FEATURES:

- Working has the highest count. But Maternity leave and Unemployed are the ones whose loans are defaulter or not approved for loan.
- Cash loans has highest count. Cash loans have more defaulters than revolving loan. (Revolving Loan- A revolving loan is considered a flexible financing tool due to its repayment and re-borrowing accommodations. Examples of revolving credit include credit cards)
- Females has most counts for applying loan. Males are more defaulters compared to female applicants.
- House/Apartment has highest counts. Applicants living in rented house and with parents having more defaulters and office apartments are more non defaulters.
- Block of flats has most counts. Specific housing are most defaulters than block of flats.
- Unaccompanied has highest counts. In name type suite mostly are of same levels but other B has more defaulters whereas family and children have less defaulters.
- Applicants having car applied for more loans. defaulters are those who doesn't have car.
- Secondary / secondary special most counts. The applicants who have lower education like lower secondary have more defaulters than with academic degree which have less defaulters.
- The day for approaching loan is of same level are almost of same count.
- Panels have highest count but Applicants having house made of wooden have more defaulters whereas applicants with monolithic (formed of a single large block of stone) have less defaulters.
- Who have realty have highest counts but who have no realty has more defaulters but they are almost on the same level.
- Married people applied for more loans. We can see in family applicants who are single or done civil marriage are high defaulters than widow and then married.
- Reg Oper Account has highest counts. Applicants having no specific account or having operational account (Operating Account means a demand deposit account maintained at the Funding Bank in Borrower's name and designated for funding that portion of each Eligible Loan not funded by a Warehousing Advance made against that Eligible Loan and for returning any excess payment from an Investor for a Pledged Loan) have more defaulters than organization account (An organization account is an account that is owned and managed by multiple Ona users.)

- Labourers applied for more loan. People with low skill labourers has highest defaulters, then drivers, labourers, cooking staff, Security staffs, waiters are followed and applicants who are accountants, high skill tech, IT staff, HR staff are less defaulters.
- Business Entity Type 3 applied for most loans. Organization having transport type 3 has highest defaulters followed by industry type 13 and type 8. less defaulters in organization in industry type 3 and trade type 4.
- People who don't have emergency in the state applied for more loans. Applicants having emergency state mode has more defaulters than who don't have.
- People owing no car has more owning realty.

2.2 NUMERICAL FEATURES:

- Applicants with 0 children are more defaulters and members with 2 also are more defaulters and both have the most count.
- In document 3 only data are given otherwise almost all data are near to 0 in heatmap. So most of the applicants are provided with document 3 and flag documents contains 0 and 1 so we can drop except doc 3.
- In 'DAYS_REGISTRATION', 'DAYS_ID_PUBLISH' not so much data is applied by applicants.
- In regions too most data are not provided by applicants.
- In contacts- FLAG_MOBIL has all data.
- external source 1 has defaulters count between (0.1-0.3), external source 2 (almost constant little varies high in 0.6), external source 3 (constant).
- In social circle no such data are provided by applicants.
- In amount required credit bureau all data provided are constant.
- Applicants who own car 57,91 years are the most defaulters and the highest count for car own age is 7 yrs.
- Hours between 9 am to 4pm most of the applicant loans are approached.
- In family members and children there is a uptrend and same with the defaulters i.e. as the family members increases defaulters increases so we can drop one of the column.

3. PREVIOUS APPLICATION:

- The shape of Previous Applications: (1670214, 37)
- Number of duplicated files are: 0

3.1. CATEGORICAL FEATURES:

- Cash loans has the highest counts. But here in previous applications we see revolving loans are having more defaulters than cash loans, consumer loan (A Consumer Loan is a loan that banks offer to customers to buy household goods and appliances and even personal devices. The most common types of consumer loans are – mortgage, auto loan, education loan, personal loan, refinance loan, and credit card.) have least defaulters, and XNA (not available list) has the highest defaulters.
- Repeater has the highest count. People who are new and repeater too has high defaulters than refreshed ones and some data are not available.
- Days for approaching loan are of same level.
- Who were not in the last application contact are higher defaulters than the who were there in the last application.
- In name suite type mostly group of people are defaulter but group B and A and spouse and partner are almost in same level and unaccompanied has the highest counts.

- Purpose for loan which is having on disclosure of goal have more defaulter and for buying a garage has the least and XAP and XNA has most counts.
- Industry type like Auto technology followed by connectivity and jewellery has high defaulters and tourism has least defaulters. Consumer electronics has most counts has some data are not available.
- Vehicles and insurance goods have high defaulters and least can be again seen in tourism and fitness. and some data are missing and mobiles has the highest counts.
- In Payment type Cash through the bank has highest counts and almost all are in same levels except where some data are not available.
- Credit and cash offices has most counts. In channel type Account payable + cash loans have high defaulters whereas countrywide and reginal or local are of same level and car dealer have the least.
- In name portfolio POS has the most counts and cards are high defaulters and cars are the least.
- Common reason for rejection is SCOFR. XAP has highest counts.
- In name yield group some data are not available. Middle has most counts whereas high is the most defaulters.
- In product type walk in has the highest defaulters. As some data are not available x-sell has most counts.
- In Product Combination Cash has the most counts and most defaulters are Card Street, card street middle, cash street high, cash x sell high.

3.2. *NUMERICAL FEATURES*:

- We can see there is a correlation between amount application and amount credit and amt goods price so we can drop one of these columns, and both Amount_Annuity and Amount_Down_Payment is necessary to work with.
- In Amount Annuity (An annuity is a fixed amount of money that you will get each year for the rest of your life. A contract between you and an insurance company that requires the insurer to make payments to you, either immediately or in the future) are defaulter in the range from (0-18000)
- In Amount down payment most defaulters are between (0-70000)
- Application submission in a day is more and has more non defaulters. We can conclude people applied before last day has defaulters more w.r.t insured on approval and we can see people who didn't applied on last day has more approval.
- DAYS_FIRST_DRAWING','DAYS_FIRST_DUE', 'DAYS_LAST_DUE_1ST_VERSION', 'DAYS_LAST_DUE', 'DAYS_TERMINATION mostly are Non Defaulters.
- Goods price between (0-5000) are mostly defaulters.
- Rate down payment (0-10000) are defaulters.
- Last application in Day and insured approached have same trend therefore can be dropped.
- In First day drawing, days last 1st version, Days termination and days last due have same trend so we can take one of them and days last due and days decision have same trend so we can take one of them.

4. BUREAU:

- The shape of Bureau : (1716428, 17)
- Number of duplicated files are: 0

4.1.CATEGORICAL FEATURES:

- The shape of Bureau : (1716428, 17)
- Number of duplicated files are: 0
- Bad debt of credit type has the highest count with respect to the Target.
- Credit Currency type 3 has the highest count with respect to the target variable.
- Loan for purchasing equipment followed by microloan has the most count.

4.2. NUMERICAL FEATURES:

- There are no such relations among the days. In Days Credit we can see records of 8 years is given but mostly defaulters are seen within 2 years.
- In Credit Days overdue mostly defaulters are below 1 year and below 0.2 count and are in negative.
- In Days Credit End date and end date fact records are given for 100 years and most defaulters are those who applied for recent years and the fact is in negative.
- In credit days update records are there for more than 100 years.
- There is no such correlation among the amount credits.

5. BUREAU - BALANCE:

- The shape of Bureau Balance: (27299925, 3)
- Number of duplicated files are: 0

5.1.CATEGORICAL FEATURES:

- Number of duplicated files are: 0 Status 4 has the most defaulters followed by 3 and 5 least has in status C.

5.2.NUMERICAL FEATURES:

- There are total 96 months records are there i.e. of 8 years and they are given in negative. And most Defaulters are within 2 years.

6. CREDIT CARD BALANCE:

- The shape of Credit Balance: (3840312, 23)
- Number of duplicated files are: 0

6.1.CATEGORICAL FEATURES:

- The Contract which are on demand has the most defaulters and the contract which has been sent proposal has least defaulters and the approved ones obviously will not have any defaulters.

6.2.NUMERICAL FEATURES:

- There is a high correlation among – AMT RECEIVABLE PRINCIPAL, AMT RECIVABLE, AMT TOTAL RECEIVABLE and between - AMT PAYMENT CURRENT, AMT PAYMENT TOTAL CURRENT, and in between AMT INST MIN REGULARITY and AMT BALANCE.
- There is a correlation among CNT DRAWINGs CURRENT and CNT DRAWINGS POS CURRENT.
- Others we can see a same trend of defaulters.

7. Instalment Payment:

- The shape of Instalment Payment: (13605401, 8)
- Number of duplicated files are: 0
- It has no Categorical features.

7.1.NUMERICAL FEATURES:

- It has no Categorical features.
- There is a correlation between AMT INSTALLMENT AND AMT PAYMENT.
- Days Install Payment is given in negative and most defaulters is in recent years like within 2 years and this is same for Days Entry of Payment.
- Install Payment Number have defaulters in 10 and 20.

8. POS CASH:

- The shape of POS Cash: (10001358, 8)
- Number of duplicated files are: 0

8.1.CATEGORICAL FEATURES:

- Amortized debt has the highest defaulters. An amortized loan is a type of loan
- with scheduled, periodic payments that are applied to both the loan's principal Amount and the interest accrued.

8.2.NUMERICAL FEATURES:

- There is a correlation between CNT INSTALLMENT and CNT INSTALLMENT FUTURE and both have a defaulter count between 0-25.