Credit EDA - Loan defaulters

By : Kumari Chanchal
Batch : DS C55

Problem statement

There are two types of risks associated with any loan request:

- 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.

Analysis of the data set has been done in Python on a Jupyter Notebook.

Analysis/Approach :

- Missing values imputation
- Variables evaluation and outlier detection
- Univariate analysis on categorical/object/character variables
- Univariate analysis on numerical variables
- Bivariate analysis on numeric variables

Analysis result :

- Majority of consumers have taken cash loans.
- Cash loan consumers are less prone to default.

1. CODE_GENDER :-

   - The majority of loans have been taken out by women.
   - Just 7% of female borrowers default, which is lower and safer than male borrowers.

Analysis result :

2. NAME_TYPE_SUITE :-

   - Unaccompanied people have paid off most of their loans, with delinquency rates of up to 8.5%, which is still okay

3. NAME_INCOME_TYPE :-

   - The safest segments are working, commercial associates and pensioners

4. NAME_EDUCATION_TYPE :-

   - Higher education is the safest segment to give the loan with a default rate of less than 5%

Analysis result :

5. NAME_FAMILY_STATUS :-

    - Married people are safe to target, as the default rate is 8%

6. NAME_HOUSING_TYPE :-

    - People having house/appartment are safe to give the loan with default rate of ~8%

7. OCCUPATION_TYPE :-

    - Low-Skill Laboreres and drivers are highest defaulters

    - Accountants are less defaulters

    - Core staff, Managers and Laborers are safer to target with a default rate of <= 7.5 to 10%

Analysis result :

5. ORGANIZATION_TYPE :-

- Transport type 3 highest defaulter

- Others, Business Entity Type 3, Self Employed are good to go with default rate around 10 %

- **Univariate numeric variables analysis :**

- Most of the loans were given for the goods price ranging between 0 to 1 ml
- Most of the loans were given for the credit amount of 0 to 1 ml
- Most of the customers are paying annuity of 0 to 50 K
- Mostly the customers have income between 0 to 1 ml

Analysis result :

- **Bivariate analysis :**

  - AMT_CREDIT and AMT_GOODS_PRICE are linearly corelated, if the AMT_CREDIT increases the defaulters are decreasing
  - People having income less than or equals to 1 ml, are more like to take loans out of which who are taking loan of less than 1.5 million, coudl turn out to be defaulters. we can target income below 1 million and loan amount greater than 1.5 million
  - People having children 1 to less than 5 are safer to give the loan
  - People who can pay the annuity of 100K are more like to get the loan and that's upto less than 2ml (safer segment)

Thank You