Credit EDA Case Study

Abhay Desai

Problem Statement - Risk Analytics in Banking and Financial Services

- This case study aims to give you an idea of applying EDA in a real business scenario.

- What we get:
  - ❑ Develop a basic understanding of risk analytics in banking and financial services
  - ❑ Understand how data is used to minimize the risk of losing money while lending to customers

- What is required?
  - ❑ results of univariate, segmented univariate, bivariate analysis, correlation etc. in business terms.

Business Understanding

- Use EDA to analyze the patterns present in the given sample data to ensure that the applicants those who are capable of repaying the loan are not rejected.

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.

Two types of scenarios in the given data :

- **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.

Business Objectives

Goal:

- **Identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as :**

  - **denying the loan,**
  - **reducing the amount of loan,**
  - **lending (to risky applicants) at a higher interest rate, etc.**

- **This will ensure that the consumers capable of repaying the loan are not rejected.**
- **Identification of such applicants using EDA is the aim of this case study**

General Process Flow

Has applied for loan previously?

No:

Yes:

Yes:

No:

Approve

Reject

Was previous loan application approved?

Find out how much was:

Loan amount , Credit amount

Annuity , Down payment

Interest rate

Any Repayment Issues:

Late payment , Missed payment

Request for increase tenure etc.

Find out:

Age, employment duration,

mobile number(reachable or not),

credit rating, income, occupation, collateral now,

Region rating , Default cases in surroundings,

Credit bureau inquiries , Loan Amount asked and Tenure.

Was it rejected because of :

low credit rating, low income,

no collateral or any other issue?

If there are any changes in those rejection parameters,

then application can be considered,

otherwise reject this time also.

Python Notebook Flow

- Import required libraries such as pandas, numpy, matplotlib and seaborn
- Load application data and previous application data
- Verify application data columns for NULL values in columns
  - If a column is having more than 40% null values, then drop that column
  - For all other columns, updated NULL to Mode / Median of that column
  - Plot various graphs for different columns
  - Plot co-relation for different columns

Plot graphs for current application…

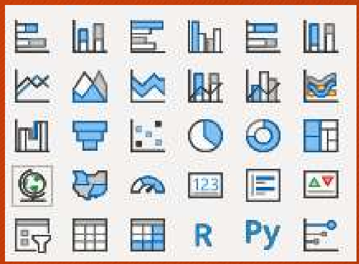
- % of defaulters Vs regular payer
- Age wise breakup of defaulters
- Loan type wise breakup in terms of defaulters and regular payer
- Gender wise breakup in terms of defaulters and regular payer
- Defaulters Vs regular payer having car and realty
- Employment /Occupation type wise breakup in terms of defaulters and regular payer
- Income wise breakup in terms of defaulters and regular payer
- Housing wise breakup in terms of defaulters and regular payer

Plot graphs for previous application…

- Loan type for previous loan applications
- Approval status for previous loan applications
- Rejection reason for previous loan applications
- Interest rate for previous loan applications
- Credit for previous loan applications
- Down payment for previous loan applications
- Number of days taken for decision for previous loan applications
- Payment type for previous loan applications

Data Cleaning and Analysis

- **Data Import:**
- **Import .CSV files provided**
- **Verify structure of imported data**



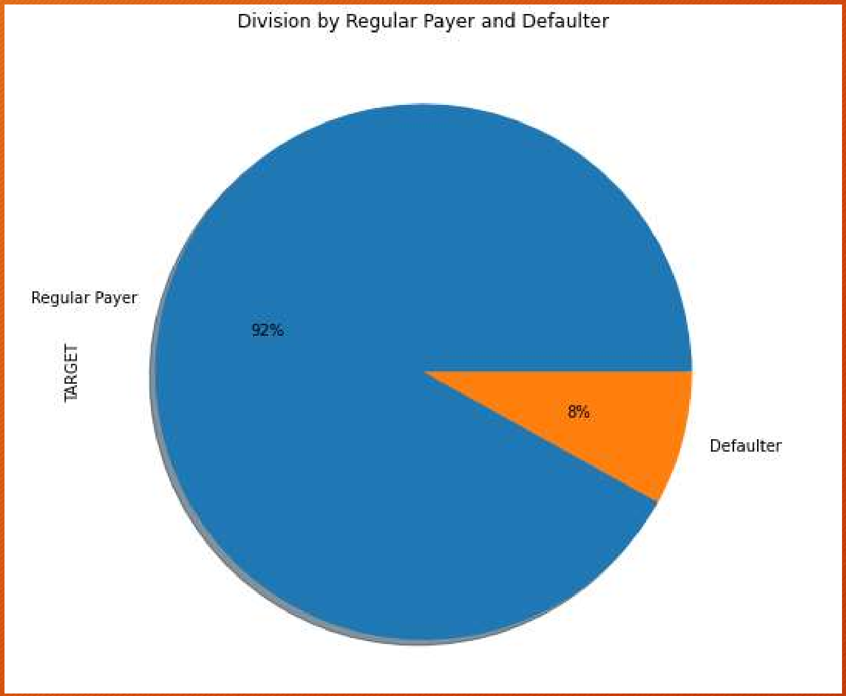
Cleansing :

- Use abs() to convert negative values to positive
- Find percentage of missing values for all columns
- Remove columns with high missing percentage and Impute default values for columns having less missing percentage
- Check datatypes of all columns and change datatype like negative age and date
- Convert number of days in years
- For numerical columns, check outliers by using boxplot & scatterplot and add observations
- Binning of continuous variables using pd.cut()

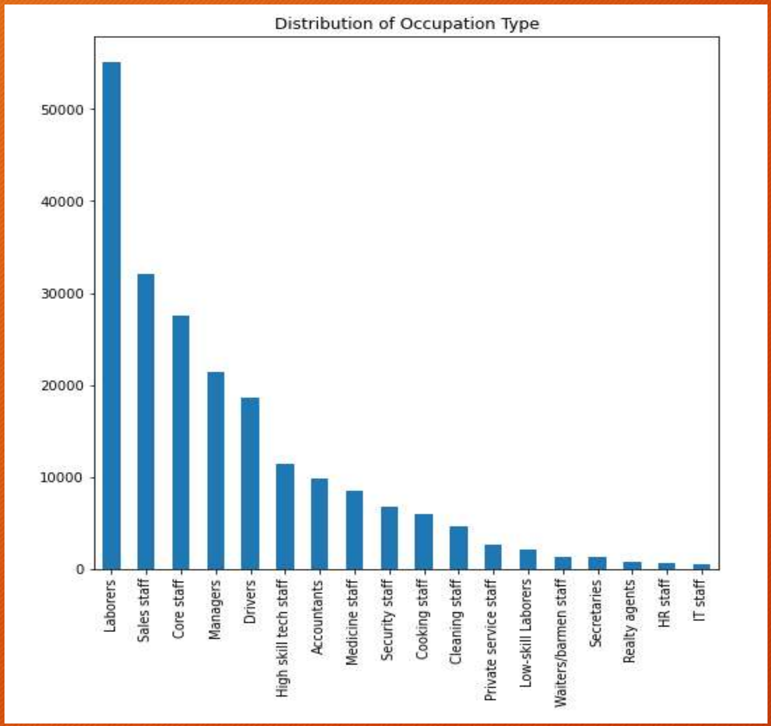
Analysis:

- check imbalance percentage using value_counts()
- divide the data into 2 datasets, for target 0 and 1
- perform univariate analysis for categorical variables- both target 0 &1
- find correlation for numerical columns for both 0 and 1 using df.corr() and seaborn heatmap
- plot 2 graphs, one for defaulters and another for non-defaulters
- Bivariate analysis such as "loan applied" Vs. "education“

Graphs & Inferences

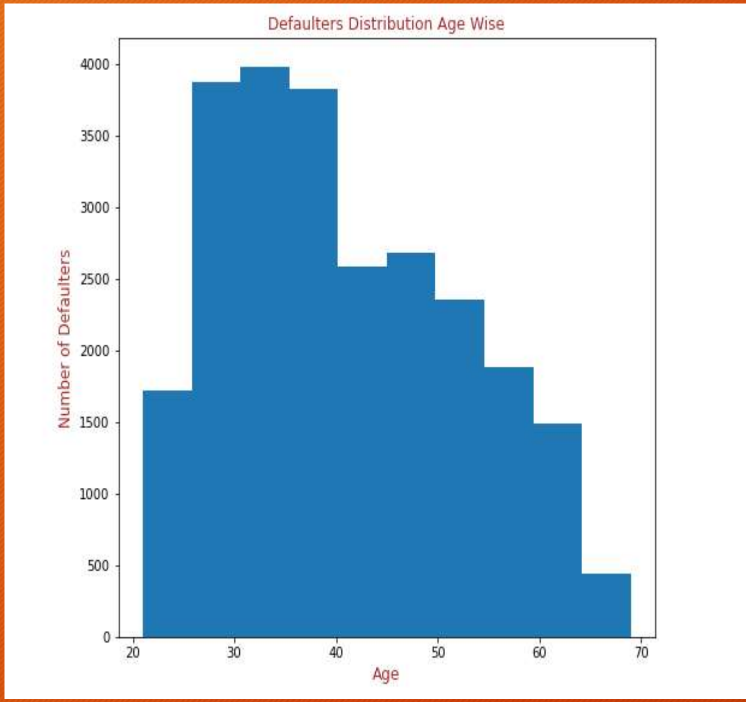


- Observations-
- 1. There are about 8% customers who are defaulters and remaining 92% are regular payers.

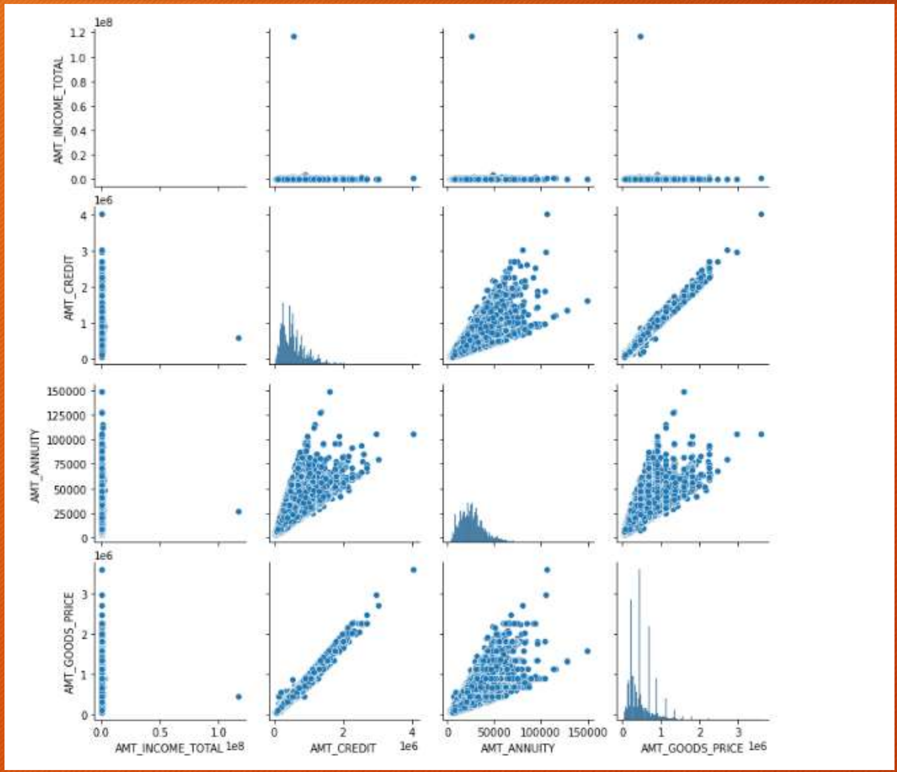


- Observations-

- 1. Majority of customers are either Laborers, Sales Staff or Core Staff.
- 2. The percentage of realty agents, HR staff and IT staff as bank customers are very less.

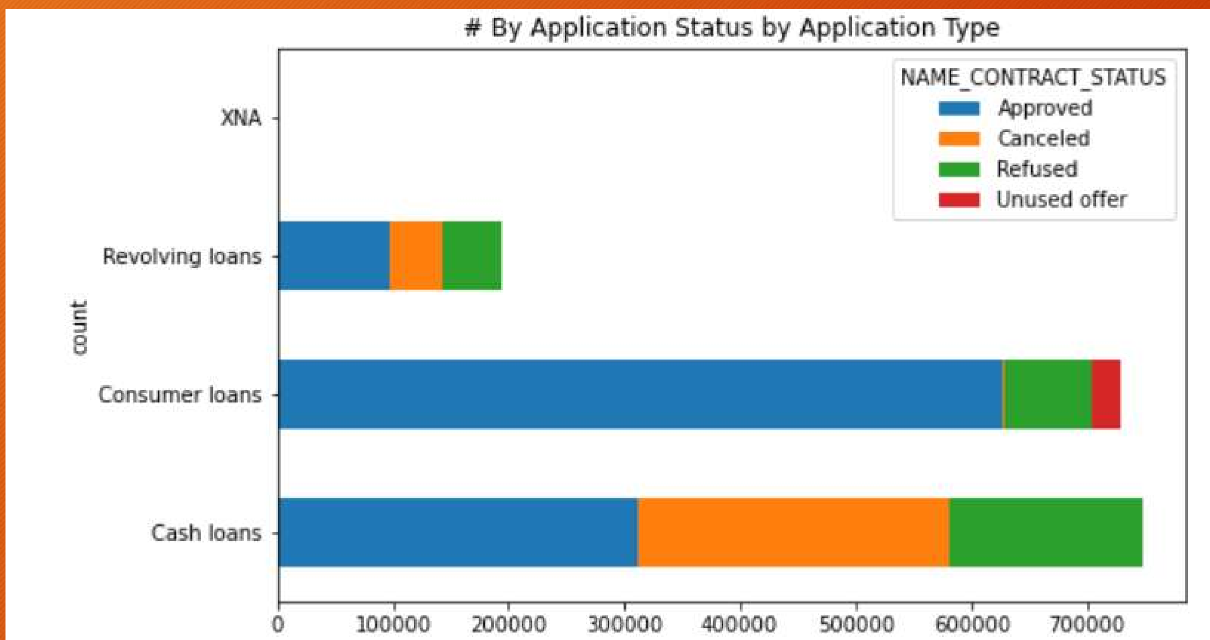


- Observations-
- 1. Majority of customers who are defaulting, are from age group of 28 to 40.
- 2. People above age group of 65 are regular payers.

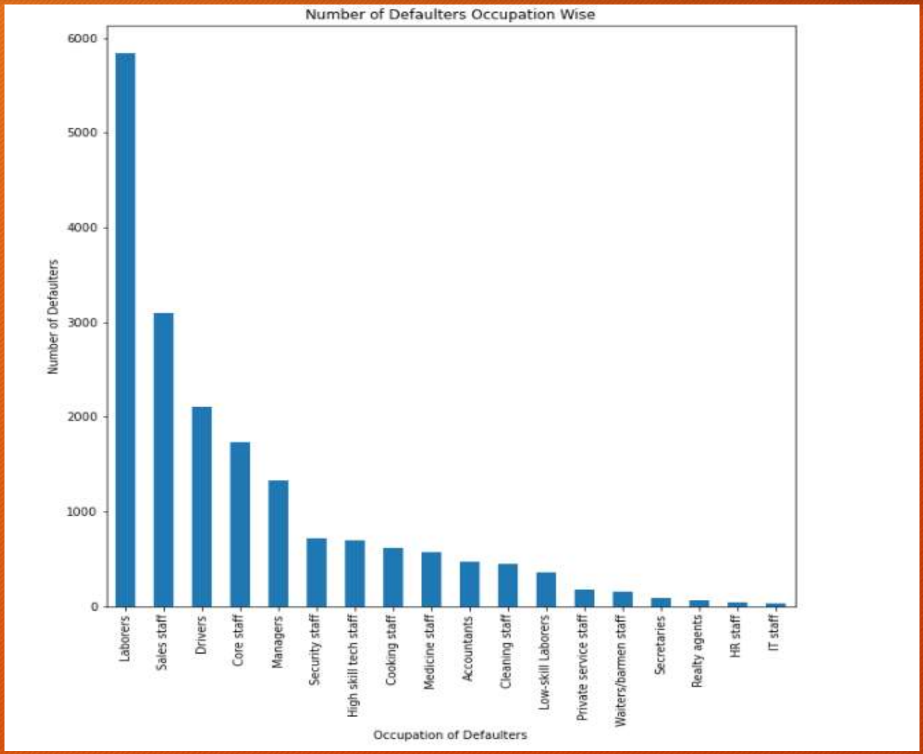


- Observations-

- 1. There is strong co-relation between "AMT_CREDIT" and 'AMT_ANNUITY".
- 2. There is also strong co-relation between "AMT_GOODS_PRICE", "AMT_CREDIT" and 'AMT_ANNUITY".
- 3. It looks like there is no so-relation between "AMT_INCOME" and "AMT_CREDIT", which needs to be thoroughly as loan amount being sanctioned should be in proportion of customer income.

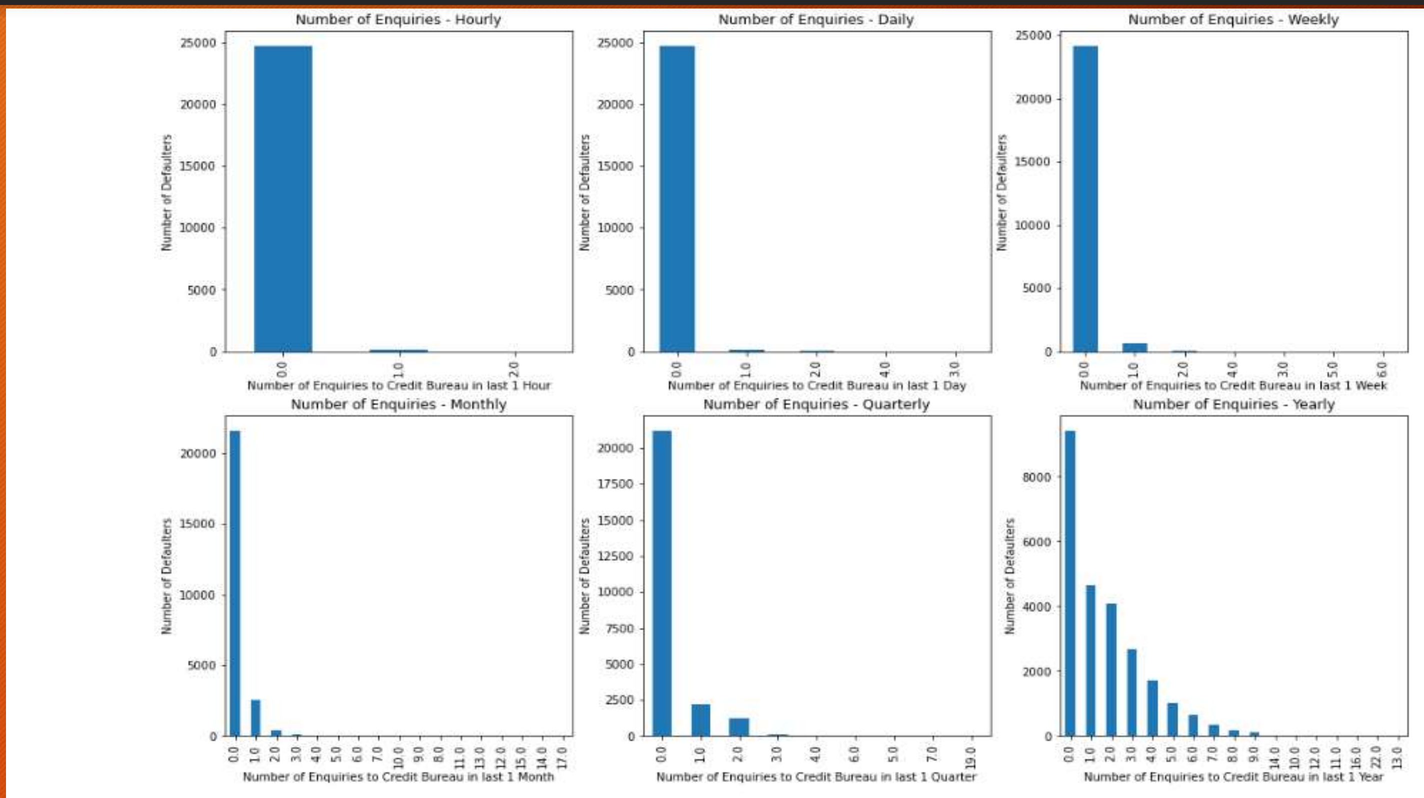


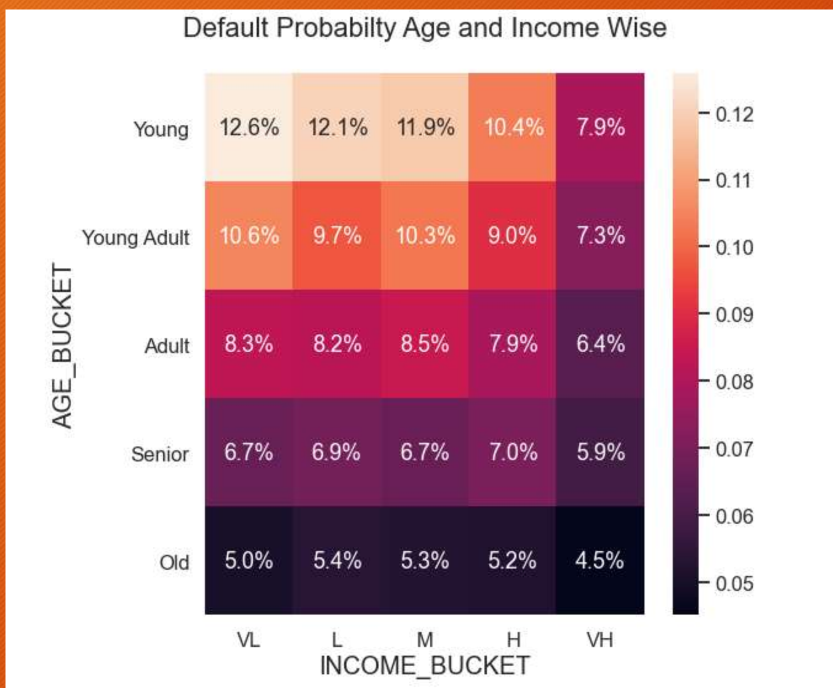
- Observations-
- 1. More “Cancellations” in cash loans.
- 2. Highest “Approvals” in Consumer loans.



- Observations-

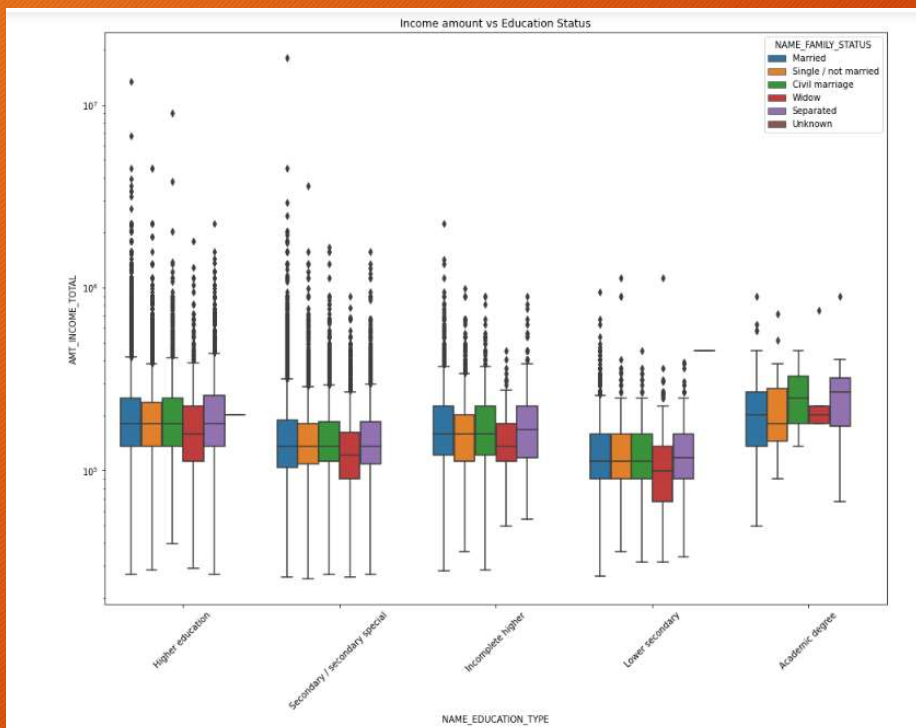
- 1. Customers having occupation as laborers, sales staff, drivers and core staff tends to default more compared to other occupations.
- 2. Customers having occupation as HR staff and IT staff tend to pay annuity regularly.


Number of Enquiries - Hourly
Number of Defaulters
Number of Enquiries to Credit Bureau in last 1 Hour
Number of Enquiries - Daily
Number of Defaulters
Number of Enquiries to Credit Bureau in last 1 Day
Number of Enquiries - Weekly
Number of Defaulters
Number of Enquiries to Credit Bureau in last 1 Week
Number of Enquiries - Monthly
Number of Defaulters
Number of Enquiries to Credit Bureau in last 1 Month
Number of Enquiries - Quarterly
Number of Defaulters
Number of Enquiries to Credit Bureau in last 1 Quarter
Number of Enquiries - Yearly
Number of Defaulters
Number of Enquiries to Credit Bureau in last 1 Year

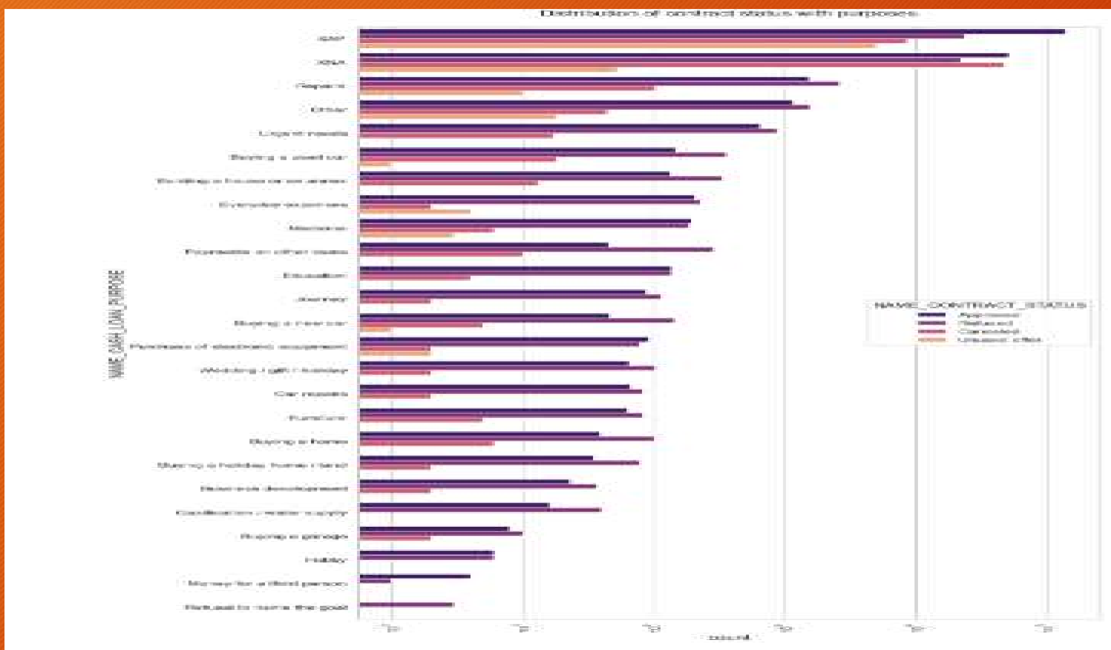


- Observations-

- 1. The probability of defaulting is higher in group of "Young" and "Young Adult" having income as "Very Low" and "Low".
- 2. Overall, the probability of defaulting is lowest in group of "Old" people.



- Observations-
- 1.For Education type 'Higher education' the income amount is mostly equal with family status.
- 2.It does contain many outliers.
- 3.Less outlier are having for Academic degree but there income amount is little higher than Higher education.
- 4.Lower secondary of civil marriage family status are have less income amount than others.



- Observations :

- 1.Most rejection of loans came from purpose 'repairs'.
- 2.For education purposes we have equal number of approves and rejection
- 3.Payign other loans and buying a new car is having significant higher rejection than approves.

Recommendations

- Banks should focus more on contract type ‘Student’ ,’pensioner’ and ‘Businessman’ with housing ‘type other than ‘Co-op apartment’ for successful payments.
- Banks should focus less on income type ‘Working’ as they are having most number of unsuccessful payments.
- Also with loan purpose ‘Repair’ is having higher number of unsuccessful payments on time.
- Get as much as clients from housing type ‘With parents’ as they are having least number of unsuccessful payments.


THANK
YOU