BFSI CASE STUDY

Submitted By:
Archana K
Amruta Patil

Stages of the case study

This case study advanced through the following stages to develop a machine learning model:

1. Problem Statement
2. Data Understanding- inspection and preparation of data
3. Exploratory Data Analysis
4. Feature Engineering
5. Handling Imbalanced Classes
6. Evaluation Metric
7. Conclusion

Problem Statement

In this case study, the objective is to build an end-to-end scoring mechanism for Home Credit, a financial institution, to assist in the decision-making process for loan applications. The goal is to predict whether a loan application should be approved or rejected based on the applicant's past behavior and application information.

Home Credit provides two datasets: application data and bureau data. The application data includes various attributes about the applicant, such as income, credit amount, family status, and education level. The bureau data contains detailed trade-level information about the applicant's credit history with other financial institutions.

Objectives:

Data Preparation:

- Clean and preprocess the application and bureau datasets.

Feature Engineering:

- Aggregate trade-level bureau data to applicant level.
- Create new features that influence loan approval.

Model Development:

- Build a classification model to predict loan approval or rejection.

Insights and Strategies:

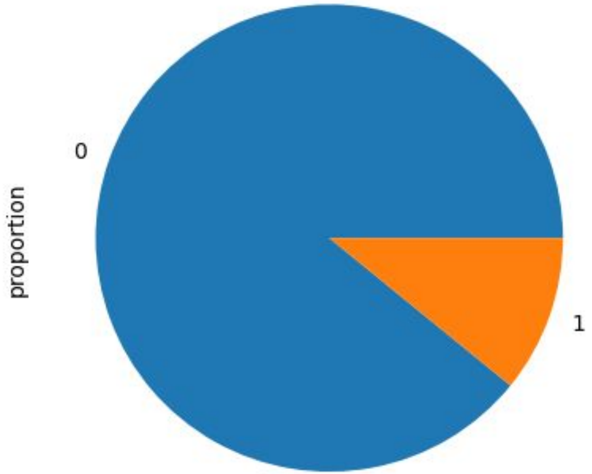
- Identify key factors affecting loan approval.
- Translate model outputs into actionable business strategies.

Key Questions:

1. How to aggregate trade-level bureau data to applicant level effectively?
2. Which factors significantly influence loan approval decisions?
3. How to build a reliable classification model for loan decisioning?
4. How to derive business insights and strategies from the model?

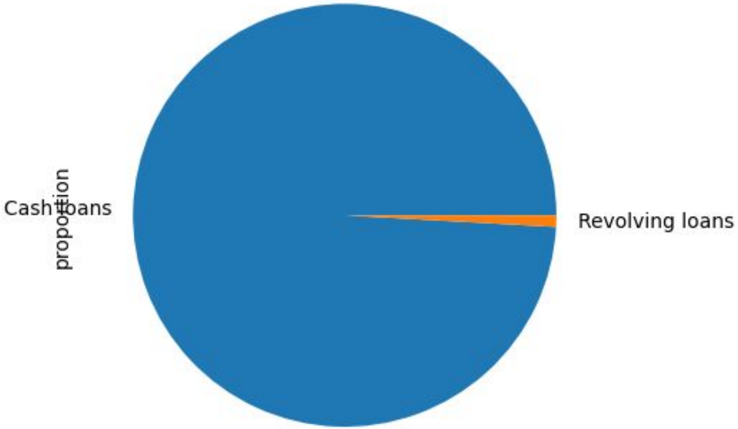
Deliverables:

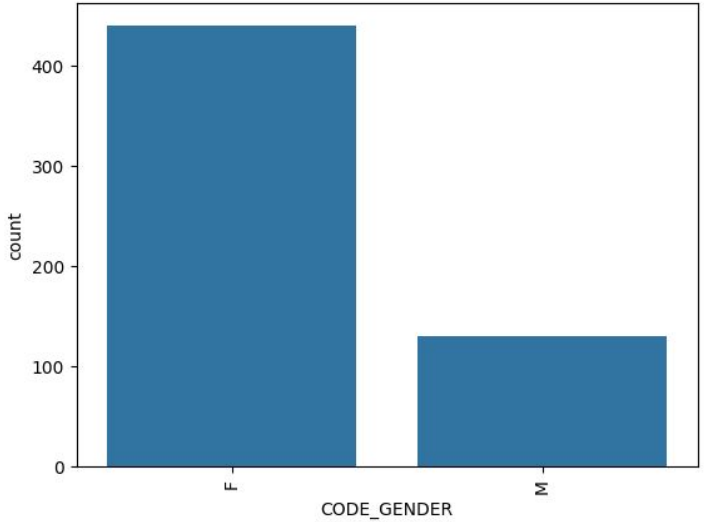
- Cleaned datasets.
- Engineered features.
- Classification model.
- Report with key factors and strategic recommendations.



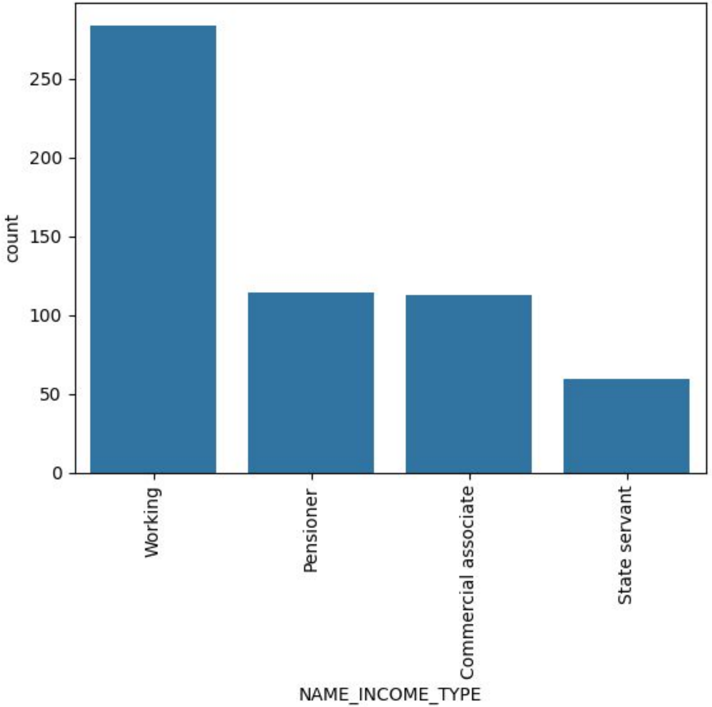
client with payment difficulty are 10 % & with other cases it is 89%

Cash loan contracts are more preferred than Revolving Loans

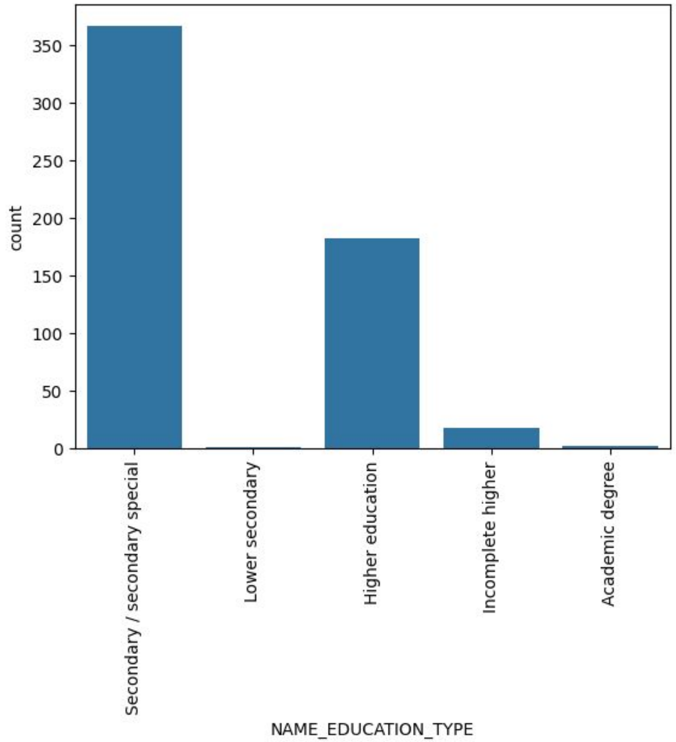




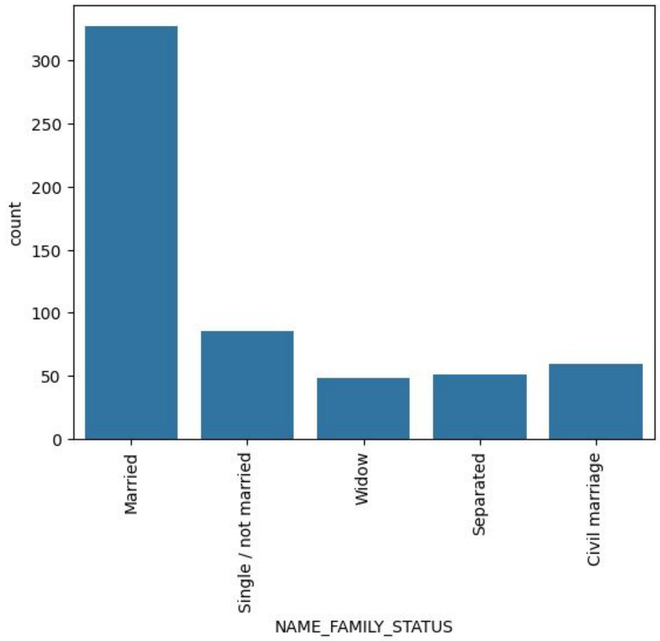
We can see here that number of Females who applied for loan are more i.e 65% ,where male who applied are 34%.



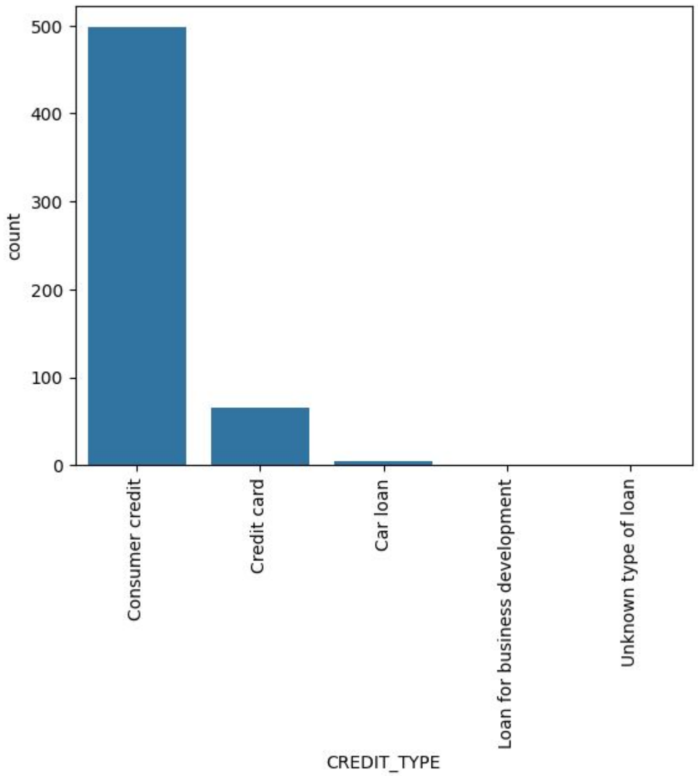
more than 50% are working



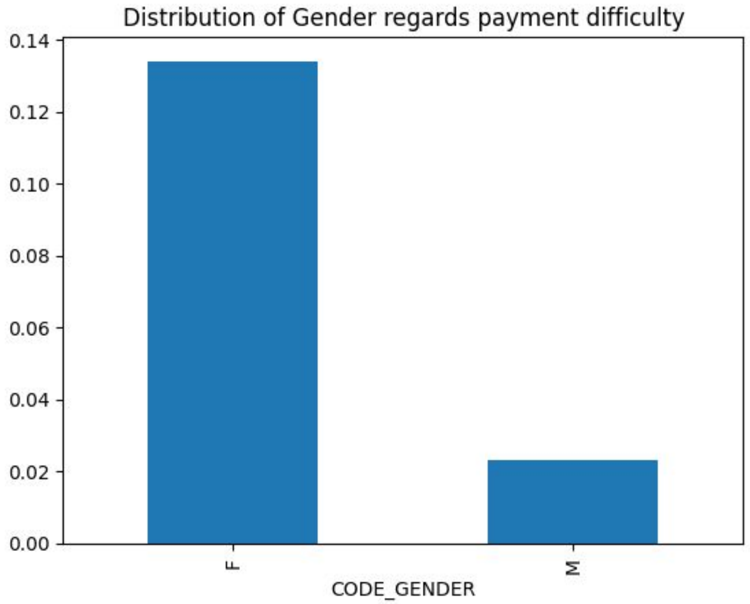
Maximum clients have taken Secondary education.



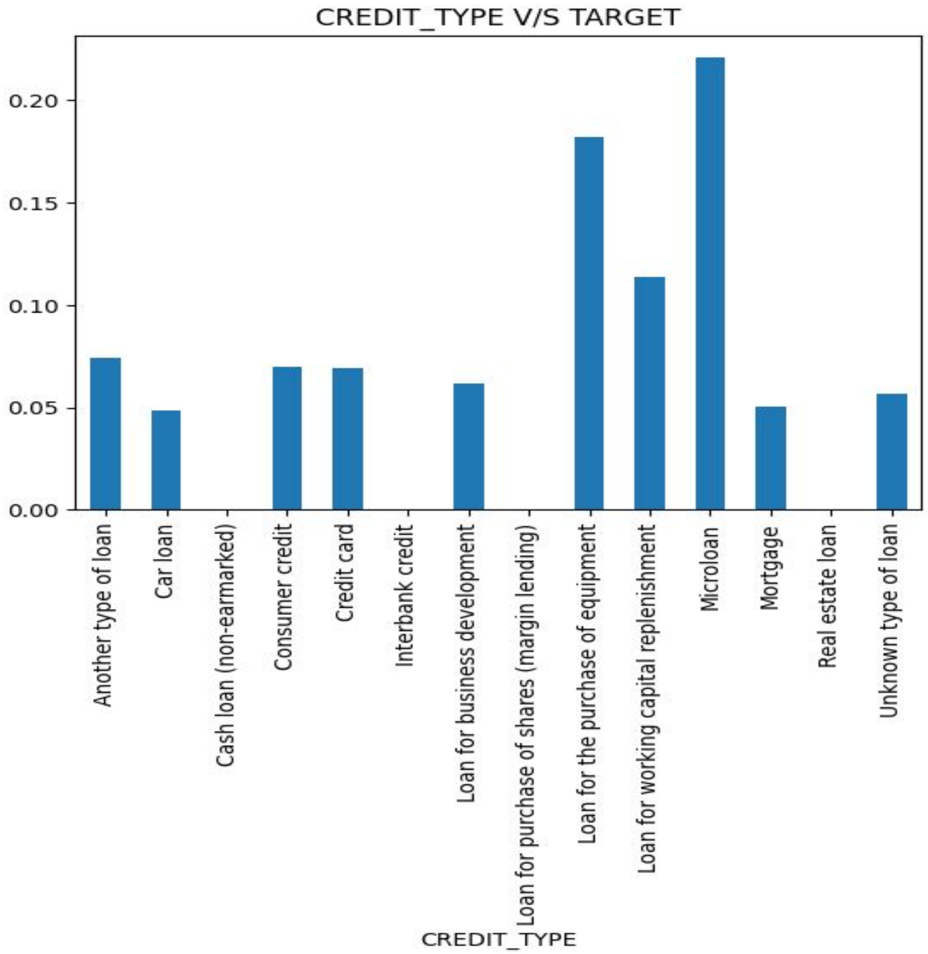
Married are in large number



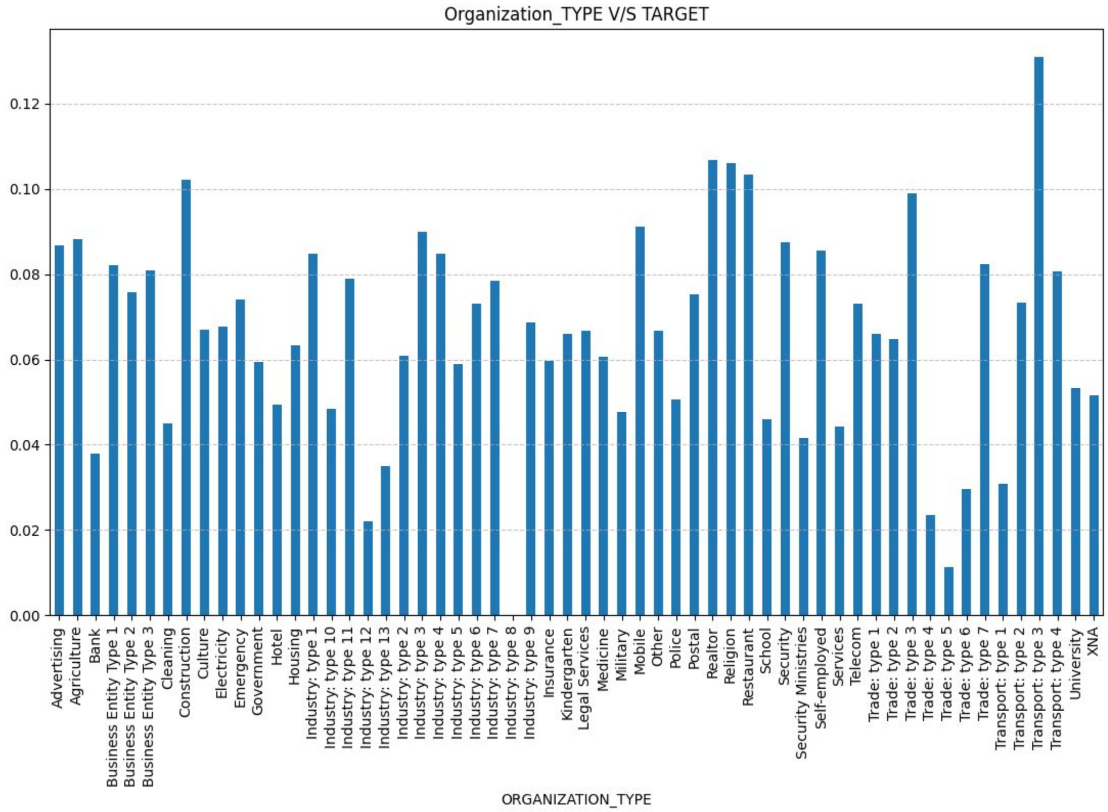
Consumer credit is taken more followed by credit card



Most loan applicants are female.



Inference:
The chart indicates that applicants with microloans, loans for working capital replenishment, and loans for the purchase of equipment are most likely to face repayment difficulties. Moderate repayment issues are seen with 'Another type of loan,' 'Consumer credit,' and 'Cash loan (non-earmarked).' In contrast, car loans, real estate loans, and mortgages show lower rates of repayment difficulties. This suggests a need for heightened risk assessment for high-risk loan categories.



The bar plot shows the distribution of the `TARGET` variable across different `ORGANIZATION_TYPE` categories. It reveals that certain organization types, such as "Transport" and "Trade", have a higher mean `TARGET`, indicating a greater likelihood of repayment difficulties. Conversely, organization types like "Electricity" and "Security" exhibit lower mean `TARGET` values, suggesting better repayment performance. This analysis highlights the variability in repayment difficulties across different organization types, aiding in identifying high-risk and low-risk categories for targeted financial strategies.

Q1.How to leverage trade level information for Credit Bureaus by aggregating trade level information to applicant level in order to capture their payment behaviour?

To capture payment behavior using trade level information, we aggregate trade-level data to the applicant level by creating features such as the number of active and closed trades, maximum overdue days, total credit amount, and most frequent currency. This feature engineering helps summarize the credit bureau data into meaningful metrics. Merging these aggregated features with application data provides a comprehensive view of each applicant's credit history. This approach enables building a robust classification model to predict loan approval decisions based on past behavior and application details.

Q2.Which application or payment behaviour factors significantly influence borrower’s behaviour on any new disbursed loan?

Significant factors influencing borrower behavior on new loans include:

1. **Credit History**: Number of active and closed trades, and maximum overdue days from credit bureau data.
2. **Income and Employment**: Total income, employment status, and type of income (e.g., working, pensioner).
3. **Demographics**: Age, education level, and marital status.
4. **Loan Characteristics**: Amount of credit applied for, type of loan (e.g., cash loan, revolving loan), and previous loan repayment behavior.

These factors provide a comprehensive understanding of an applicant's financial stability and repayment capability, aiding in accurate loan approval decisions.

Q3.how to leverage them in the form of a model which can be used for decisioning?

To leverage significant factors in a model for decision-making:

1. **Feature Engineering**: Combine and transform trade-level and application data, creating features like active/closed trade counts and overdue days.

2. **Data Preprocessing**: Handle missing values, encode categorical variables, and normalize data.

3. **Model Selection**: Choose and train classification models (e.g., Logistic Regression, Random Forest). Validate using cross-validation.

4. **Model Evaluation**: Assess with metrics like accuracy and recall. Optimize with hyperparameter tuning.

5. **Deployment**: Implement the model in decision systems and continuously monitor and update it to maintain performance.

This approach ensures accurate, data-driven loan approval decisions.

Q4.how to translate the model output into strategies and business insights for the bank?

To translate model output into strategies and business insights for the bank:

1. **Risk Segmentation**: Classify applicants into risk segments (low, medium, high) and tailor strategies accordingly.

2. **Resource Allocation**: Focus resources on monitoring high-risk borrowers and automate approvals for low-risk applicants.

3. **Policy Adjustments**: Refine lending policies and set dynamic credit limits based on applicant risk profiles.

4. **Targeted Marketing**: Design targeted marketing campaigns and personalized offers based on risk profiles.

5. **Performance Monitoring**: Continuously track the model's impact on loan performance and update the model with new data.

These steps help improve decision-making, optimize resource use, and enhance loan portfolio performance.

Summary:

The age group 30-40 has the highest number of applicants, followed by the 40-50 age group.

- The minimum age of applicants is 21 years, and the maximum age is 70 years.

- A significant number of applicants applied unaccompanied, meaning they came alone.

- The majority of applicants have completed secondary education.

- More than 50% of the applicants are currently working.

- 18% of the applicants are pensioners.

- A large number of applicants are married.

- The predominant loan type is cash loans.

- Most loan applicants are female.

- The majority of credit bureau reports indicate a closed status for credits.

- Consumer credit is the most common credit type reported by the credit bureau.

-We can see that around 4.5L applicants Occupation Type is missing and most of the applicants who has borrowed loans are laborers(Daily wages working class applicant), Core Staff, sales staff and Managers.