Exploratory Data Analysis on Loan Default Risk

By Ng Shao Qian

Report submitted in fulfilment of the module ***Hands-On Exploratory Data Analysis With Python*** under ***Certified Associate in Data Analytics*** on TalentLabs.

Overview and Background

This report presents an exploratory data analysis (EDA) of loan default risk, providing insights into factors influencing clients' ability to meet their loan repayment obligations. Risk analytics in banking and financial services plays a crucial role in minimizing financial losses while ensuring loans are provided to clients who can repay them. The goal of this analysis is to leverage data to identify patterns that indicate a higher likelihood of default, allowing companies to make informed decisions regarding loan approvals.

The case study covers two datasets: one containing details of current loan applications (`application_data.csv`) and another capturing previous loan applications (`previous_application.csv`). The analysis aims to highlight key relationships between various customer and loan attributes and their association with repayment difficulties.

Business Understanding

In the financial services industry, lending companies face challenges when approving loans for individuals with insufficient or no credit history. This often leads to a risk of defaults, where clients fail to meet their repayment obligations. The company in this scenario specializes in lending to urban customers, and the primary goal is to use data to assess which applicants are likely to repay the loan.

When a company receives a loan application, it must make a critical decision: approve or reject the loan. There are two main risks associated with this decision:

1. Rejecting a loan from an applicant who is likely to repay it results in lost business.
2. Approving a loan for a client who is likely to default may lead to financial loss.

The data provided includes two types of clients:

- **Clients with payment difficulties**: those who have missed payments on one or more installments.
- **All other cases**: clients who have consistently made their payments on time.

The company also categorizes loan applications into four statuses:

1. **Approved**: The loan is approved by the company.
2. **Cancelled**: The client canceled the application during the approval process.
3. **Refused**: The loan was rejected due to the client not meeting certain criteria.
4. **Unused Offer**: The loan was offered but not used by the client at a later stage.

The objective is to analyze the patterns in the data to help the company make informed loan approval decisions, mitigate risk, and optimize lending strategies.

Data Understanding

This analysis uses two main datasets:

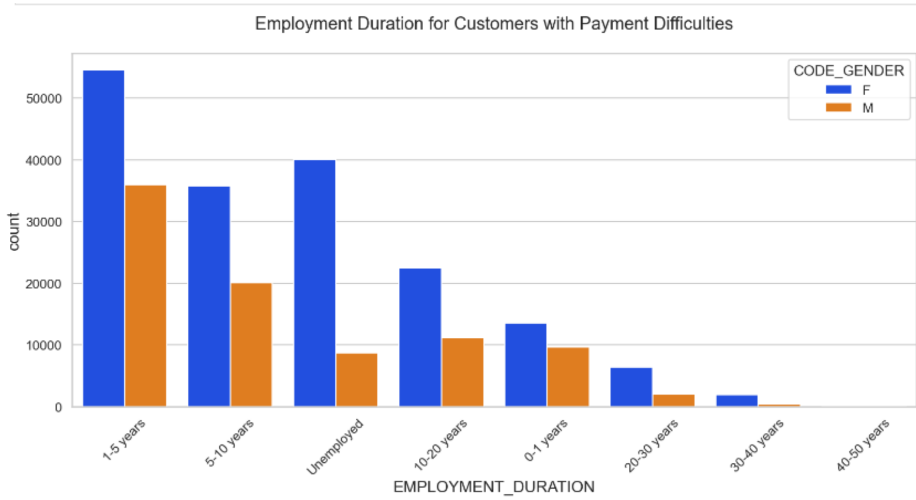
1. **`application_data.csv`**: Contains information about current loan applicants and whether they have payment difficulties.
2. **`previous_application.csv`**: Contains historical data about clients' previous loan applications, including the status of each application (approved, canceled, refused, or unused offer).
3. **`columns_description.csv`**: A data dictionary providing detailed explanations of the variables in the datasets.

Exploratory Data Analysis (EDA)

Detailed EDA is done on **`EDA EDA_on_Loan_Default_Risk.ipynb`**. Below are parts of the analysis.

Distribution of Employment Duration

Clients employed for 1-5 years are more likely to face payment difficulties than those with other employment durations. This might indicate some financial instability during this period.

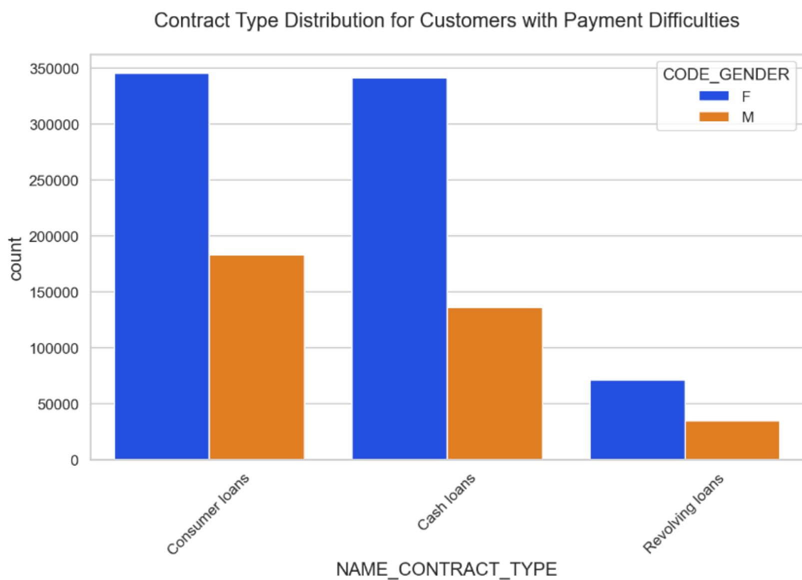


Gender

Males are marginally more likely than females to face payment difficulties.

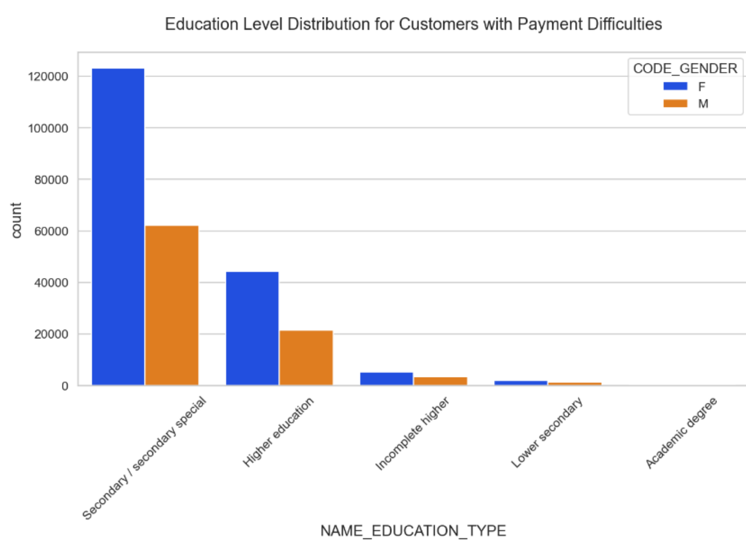
Proportion of Contract Types

Clients with revolving loans are marginally more likely to default than those with other loan types, possibly because these loans often involve revolving credit, which may lead to overborrowing. Clients with consumer loans have a small negative correlation with default, possibly reflecting that these loans are generally smaller or more manageable compared to revolving loans.



Proportion of Education Types

Clients with secondary-level education have a slight positive correlation with payment difficulties, suggesting that education level could be a factor in financial challenges. Clients with higher education are less likely to default, likely due to better employment opportunities and financial security.

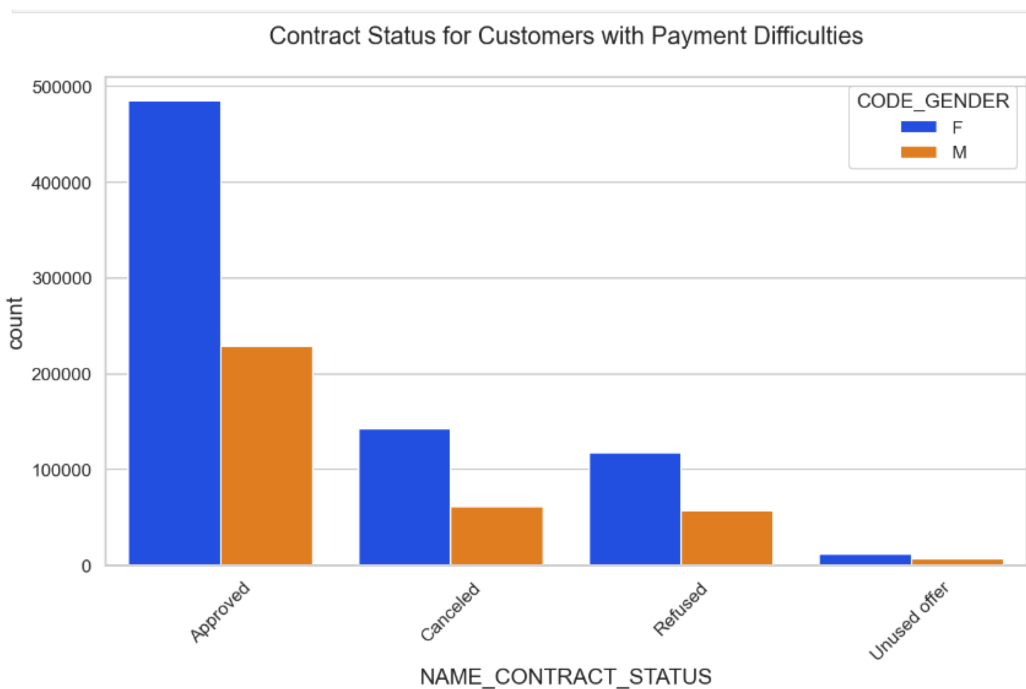


Client Region Rating

Higher region ratings correlate with a slight increase in payment difficulties, potentially indicating that living in certain regions may be linked with higher default rates.

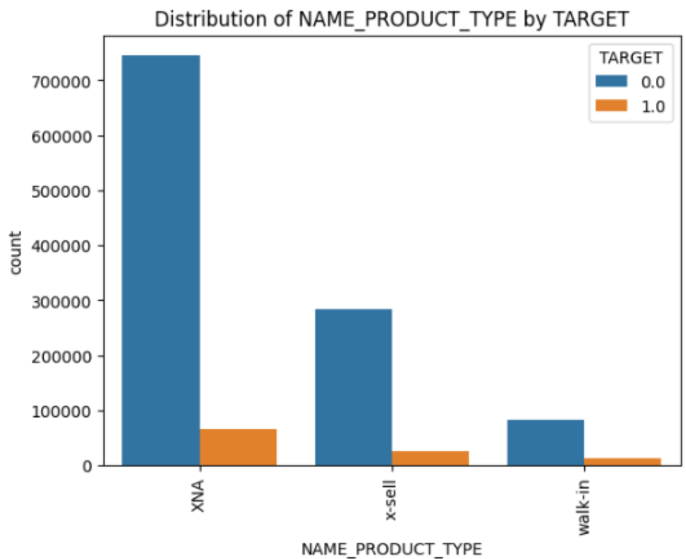
Contract status

Clients with previously refused contracts are slightly more likely to face payment difficulties, indicating potential credit risk for individuals who have faced rejection in past applications.



Product Types

Clients who used "walk-in" product types for their previous loan applications show a small positive correlation with payment difficulties, possibly indicating that these clients may not have used sophisticated loan services, leading to financial risk.

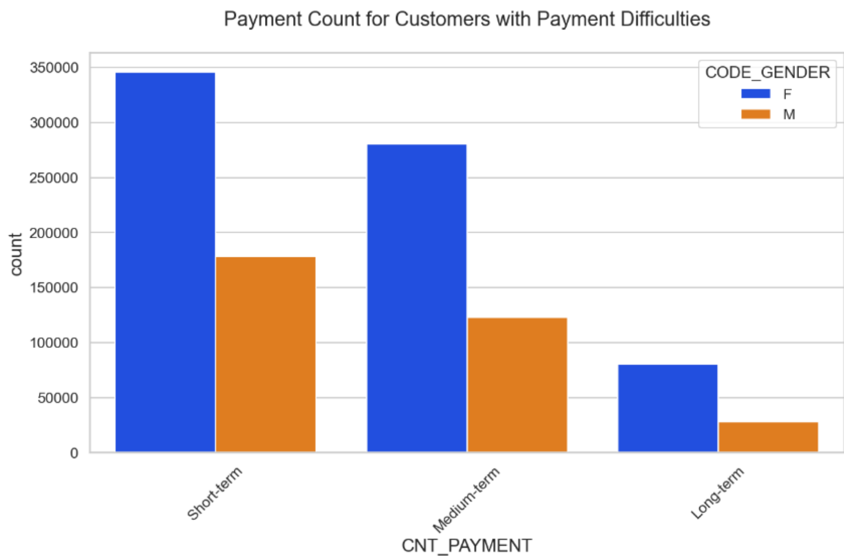


Goods category

Clients with goods categorized under "XNA" (likely missing or unknown information) show a slight positive correlation with payment difficulties, possibly due to incomplete application information. In contrast, purchases categorized as "Jewelry", “Weapons”, “Direct sales”, “Animals” have almost no correlation with default risk, suggesting that this category doesn't significantly influence financial stability.

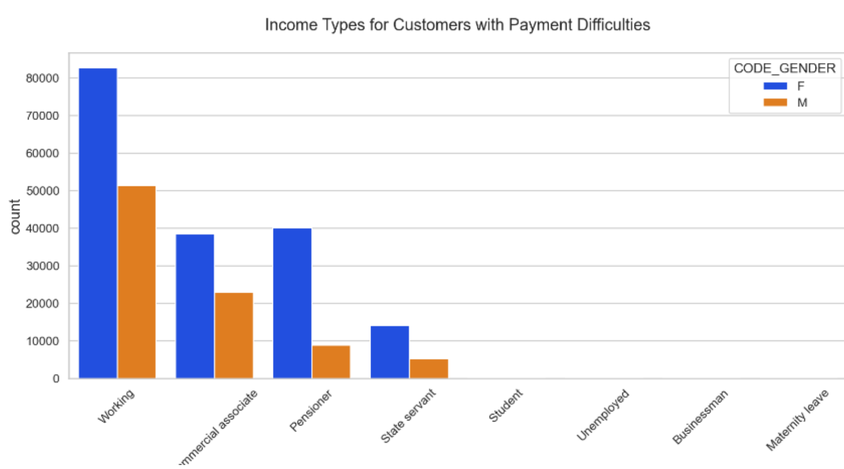
Payment term

Clients with long-term payments have a small increase in default risk, possibly due to the longer commitment required for such loans.



Income type

Pensioners are less likely to suffer from payment difficulties, which could reflect greater financial security from retirement income. However, working clients have a positive correlation with payment difficulties, possibly because salaried or wage-earning clients experience variable income or financial instability compared to other income types (e.g., business owners).



Research Questions and Insights

1. **How does employment duration affect the likelihood of default?**

    - **Finding**: Employment duration between 1-5 years shows a positive correlation with loan defaults. Clients in this group face more financial instability, possibly due to less secure job positions.
    - **Recommendation**: Implement stricter credit assessments for applicants with short-term employment and consider offering financial planning resources to mitigate risk.

2. **Does education level influence default rates?**

    - **Finding**: Higher education levels are negatively correlated with loan defaults. Clients with advanced education are more likely to repay their loans, possibly due to better employment opportunities and financial stability.
    - **Recommendation**: Offer more favorable loan terms to clients with higher education levels and provide financial literacy programs for lower-educated clients.

3. **Do external credit scores (EXT_SOURCE_2 and EXT_SOURCE_3) predict defaults?**

    - **Finding**: Strong negative correlations with external credit scores suggest that clients with high credit scores are less likely to default. These scores are among the strongest predictors of repayment ability.
    - **Recommendation**: Use external credit scores as a primary factor in loan approval decisions. Consider higher interest rates for lower-scoring clients or offering lower-risk loan products.

4. **How does income type influence loan default rates?**

    - **Finding**: Pensioners are less likely to face payment difficulties, reflecting greater financial security from retirement income. In contrast, working clients show a positive correlation with defaults, possibly due to variable income or financial instability compared to other income types (e.g., business owners).
    - **Recommendation**: Offer more favorable loan terms to pensioners, given their lower risk, and consider providing more structured loan options or higher interest rates to wage earners to account for income variability.

5. **Do contract types affect the likelihood of default?**

    - **Finding**: Clients with revolving loans are slightly more likely to default, possibly due to the nature of revolving credit leading to overborrowing. Conversely, consumer loans show a weak negative correlation with defaults, reflecting that these loans are generally smaller and easier to manage.
    - **Recommendation**: Implement stricter approval criteria or higher interest rates for revolving loans to mitigate default risk, while offering consumer loans with more flexible terms to encourage responsible borrowing.

Conclusion

This EDA has revealed several important factors influencing loan default risks. Key insights include the reduced risk for pensioners and clients with consumer loans, as well as the increased risk for clients with revolving loans or shorter employment durations. Region ratings and education levels also play a role, with higher region ratings and lower education levels being linked to greater default risk.

The company can leverage these findings to design risk-based loan products and fine-tune approval processes. For example:

- **Income Type-Based Products**: Favorable terms can be extended to pensioners, while working clients may benefit from structured loan products that account for income variability.
- **Contract Type Management**: Stricter criteria for revolving loans can help manage the increased risk, and consumer loans can be made more accessible to clients with stronger repayment abilities.
- **Region-Based Adjustments**: Introducing region-based risk scoring can help better predict and mitigate potential default risks based on location.

These actions will not only reduce financial losses but also optimize loan approval strategies, ensuring the business balances growth with effective risk management.