# Analyzing Business Problem

# Specialization: Risk Analysis

## 1. Task requirements

In the world of e-commerce, understanding how much time it takes for users to make a purchase on our website is vital. This analysis helps us see how efficient our website is and identify areas for improvement.

The final Basel III framework will be implemented on 1 of January 2024. Your manager asks you to calculate credit risk RWA and capital requirement for the bank's portfolio.

You must use data from the file Mortgage default, which contains loans with a status indicator: defaulted(bad) or non-defaulted. Here is short description of data columns:

BAD: 1 = applicant defaulted on loan or seriously delinquent; 0 = applicant paid loan.
MORTDUE: Amount due on existing mortgage.
VALUE: Value of current property.
JOB: Occupational categories.
YOJ: Years at present job.
DEROG: Number of major derogatory reports.
DELINQ: Number of delinquent credit lines.
CLAGE: Age of oldest credit line in months.
NINQ: Number of recent credit inquiries.
CLNO: Number of credit lines.
DEBTINC: Debt-to-income ratio.

Detailed columns description is provided here.

As you have practiced in your hands on task, now you must calculate credit risk RWA for the given data under both: standardized (follow the regulation, as in the hands-on task) and IRB approach and compare the results.

For the PD model, the model's goodness of fit will not be evaluated. All given loans in the dataset are mortgage loans. If you have limitations in google spreadsheets to calculate logistic regression with a full dataset, use less. data or choose another tool. For the lines with missing collateral values, use default LTV = 0.45.

Create effective visualizations, by using a tool that you are familiar with such as Google Spreadsheets, Tableau, Looker Studio, or any other suitable data visualization tool, and briefly comment your findings.

Evaluation criteria for a Graded project submission:

- Correct columns identified to calculate risk parameters, visual analyses.

- Correct calculation logic of risk parameters.
- Argumentation of chosen steps in your project.
- Visualization is clear and communicates the message/answer.
- Analysis, findings and main points clearly structured.
- During a task review, you may get asked questions that test your understanding of covered topics.

## 2. Objectives

For the business to understand default risk applying different risk assessment approaches I had several objectives.

- Calculate credit risk weight using both standartized and internal ratings based (IRB approach)
- Identify the most / least risky assets in the mortage portfolio
- Compare to approaches and provide recommendation for usage and decreasing RWA

By achieving them we will have a better understanding of our customers and what approach is better for our financial institution.

## 3. Prepare & Process

In this part I went through the data source and checked what kind of data I have.

**Data cleaning**

Missing values with 2% to 5% of total were cleaned, and missing values in the column with more than 8% were replaced by median.

**Managing outliers**

To estimate the outlier, the IQR method was used to count percentage of outliers. Since outliers' percentage for numerical data (0,1% to 7%) and they represent genuine variations in the data, I had left them in the dataset. Removing such a small percentage of data points may lead to a loss of valuable information.

| | BAD | MORTDUE | VALUE | JOB | YOJ | DEROG | DELINQ | CLAGE | NINQ | CLNO | DEBTINC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Max | 1.00 | 399,550.00 | 855,909.00 | 6.00 | 41.00 | 10.00 | 15.00 | 1,168.23 | 17.00 | 71.00 | 203.31 |
| Min | 0.00 | 2,063.00 | 8,000.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.52 |
| Mean | 0.20 | 72,809.13 | 101,776.05 | 2.98 | 8.92 | 0.25 | 0.45 | 179.77 | 1.19 | 1.19 | 21.30 |
| Standard dev. | 0.40 | 44,457.61 | 57,385.78 | 2.06 | 7.57 | 0.85 | 1.13 | 85.81 | 1.73 | 10.14 | 8.60 |
| Median | 0.00 | 65,019.00 | 89,235.50 | 2.00 | 7.00 | 0.00 | 0.00 | 173.47 | 1.00 | 20.00 | 34.82 |
| Missing | 0.00 | 469.00 | 112.00 | 279.00 | 515.00 | 708.00 | 580.00 | 308.00 | 510.00 | 222.00 | 1,267.00 |
| Missing % from total | 0% | 8% | 2% | 5% | 9% | 12% | 10% | 5% | 9% | 4% | 21% |
| | | | | | | | | | | | |
| IQR method for outliner calculation | | 48152.5 | 66000 | 1 | 3 | 0 | 0 | 114.9573955 | 0 | 15 | 29.34047633 |
| | | 89272.5 | 119623.5 | 5 | 13 | 0 | 0 | 232.324439 | 2 | 27 | 39.10820064 |
| IQR | | 41120 | 53623.5 | 4 | 10 | 0 | 0 | 117.3670435 | 2 | 12 | 9.767724308 |
| Lower bond | | -13527.5 | -14435.25 | -5 | -12 | 0 | 0 | -61.09316974 | -3 | -3 | 14.68888987 |
| Upper bond | | 150952.5 | 200058.75 | 11 | 28 | 0 | 0 | 408.3750043 | 5 | 45 | 53.7597871 |
| | | | | | | | | | | | |
| Number of outliers | | 235 | 293 | 0 | 90 | 664 | 1119 | 44 | 146 | 177 | 69 |
| % of total | | 4% | 5% | 0% | 2% | 12% | 21% | 1% | 3% | 3% | 1% |

## 4. Analyze & Share

**Calculate credit risk weight using both standartized and internal ratings based (IRB approach)**

Firstly, we will use the standardized approach. For the standardized approach, the latest regulatory requirement – Basel III (in EU, it is transformed into CRR IV) has a more granular and revised method than previous Basels. In the standardized approach the risk weights are given, so you just need to classify loans and apply risk weights.

**Standardized Approach**

1. **Exposure at Default (EAD):**
   I used the mortgage due value from the dataset as the EAD for each mortgage.

2. **Loan-to-Value (LTV) Ratio Calculation:**
   I calculated the LTV for each mortgage using the formula:

$$\text{LTV} = \frac{\text{Mortgage Due Value}}{\text{Mortgage Value}}$$

   This calculation provided the LTV ratio, which was then used to determine the risk weight.

3. **Risk Weight (RW) Assignment:**
   - Based on the LTV, I assigned the appropriate risk weight according to the following scale:
     LTV≤50%LTV≤50%: RW = 20%
     50%<LTV≤60%50%<LTV≤60%: RW = 25%
     60%<LTV≤80%60%<LTV≤80%: RW = 30%
     80%<LTV≤90%80%<LTV≤90%: RW = 40%
     90%<LTV≤100%90%<LTV≤100%: RW = 50%
     LTV>100%LTV>100%: RW = 70%

4. **Calculation of Risk-Weighted Assets (RWA):**
   I calculated the RWA for each mortgage using the formula:

$$\text{RWA} = \text{EAD} \times \text{RW}$$

   Here, EAD is the mortgage due value from the dataset.

5. **Capital Requirement Calculation:**
   The total capital requirement was calculated as 8% of the total RWA:

$$\text{Capital Requirement} = 0.08 \times \sum \text{RWA}$$

**Internal Ratings-Based (IRB) Approach**

1. **Variable Correlation and Multicollinearity Check:**

   **Correlation with Target Variable:**

   - I assessed the correlation between each independent variable and the target variable, the probability of default (PD).
   - None of the independent variables exhibited a strong correlation (absolute value close to 1) with the target variable, suggesting that no single variable strongly predicts the target variable on its own.

   **Multicollinearity:**

   - I identified pairs of variables with high correlation coefficients, such as "MORTDUE_updated" and "VALUE" (correlation coefficient ~0.85).
   - Variables with strong multicollinearity were excluded from the model to prevent redundancy and ensure model stability.

2. **Logistic Regression Model:**
   - I developed a logistic regression model to estimate the probability of default (PD) for each mortgage.
   - Variables included in the model were those that showed significant correlation with the target variable but had negligible mutual multicollinearity.
3. **Risk Components Calculation:**
   - **Probability of Default (PD):** Estimated from the logistic regression model.
   - **Loss Given Default (LGD):** Assumed to be 0.45.
   - **Exposure at Default (EAD):** Taken from the standardized approach (mortgage due value).
4. **Calculation of Risk-Weighted Assets (RWA):**

   I calculated the RWA for each mortgage using the IRB formula:

$$\text{RWA} = \text{EAD} \times \text{RW}$$

   Where RW is determined by:

$$\text{RW} = \left( \frac{1.06 \times \text{PD} \times (1 - \text{PD})}{0.12} \times \text{LGD} \times (\text{Maturity factor}) \right)$$

5. **Capital Requirement Calculation:**
   - Similar to the standardized approach, I calculated the total capital requirement:

$$Correlation = R = \ 0.12 \cdot \frac{\left(1 - e^{-50 \cdot PD}\right)}{\left(1 - e^{-50}\right)} + 0.24 \cdot \left(1 - \frac{\left(1 - e^{-50 \cdot PD}\right)}{\left(1 - e^{-50}\right)}\right)$$

$$Maturity\ adjustment = b = \left[0.11852 - 0.05478 \cdot \ln\left(PD\right)\right]^2$$

$$Capital\ requirement = K = \left[LGD \cdot N\left[\frac{G\left(PD\right)}{\sqrt{\left(1-R\right)}} + \sqrt{\frac{R}{1-R}} \cdot G\left(0.999\right)\right] - PD \cdot LGD\right] \cdot \frac{\left(1 + \left(M - 2.5\right) \cdot b\right)}{\left(1 - 1.5 \cdot b\right)}$$

$$RWA = K \cdot 12.5 \cdot EAD$$

By following these steps, I was able to determine the exposure at default, risk weights, risk-weighted assets, and capital requirements for both the standardized and IRB approaches. The standardized approach relied on predefined risk weights based on LTV ratios, while the IRB approach involved a more detailed analysis using a logistic regression model to estimate the probability of default and calculate corresponding risk parameters.

**Compare to approaches and provide recommendation for usage and decreasing RWA**

Comparing of standartized and IRB approaches I have noticed such differences:

**Capital Requirement:**
The capital requirement under the IRB Approach is significantly higher ($37.3M) compared to the Standardized Approach ($9.2M). This indicates that the IRB Approach assigns a much higher capital buffer for non-defaulted loans based on its risk assessment.

**Risk Weight (RW):**
The RW under the IRB Approach is also notably higher (141.44%) compared to the Standardized Approach (34.75%). This suggests that the IRB Approach assigns a higher risk weight to non-defaulted loans, reflecting a higher perceived risk compared to the Standardized Approach.

**Total RWA:**
The total Risk-Weighted Assets (RWA) under the IRB Approach are substantially higher ($466.8M) compared to the Standardized Approach ($114.7M). This indicates that the IRB Approach results in a larger capital requirement due to a higher assessment of risk associated with the loan portfolio.

**Identify the most / least risky assets in the mortage portfolio**

To accurately assess the risk levels within a credit portfolio, it is essential to analyze the default probabilities across various borrower characteristics. The descriptive analysis below offers insights into how occupational categories, the number of delinquent reports, and tenure at the current job influence the risk of mortgage default.

**Occupational Categories**

Different occupational categories exhibit varying levels of default risk, which can be quantified by the ratio of defaulted Mortgage Due (MortDue) amounts to non-defaulted MortDue amounts. Here's the breakdown:

- Sales (0.483) and Self-employed (0.323): These categories have the highest defaulted to non-defaulted MortDue ratios, indicating a significantly higher risk of default.
- Other (0.234) and Managerial (0.237): These occupations fall into a moderate risk category, with default ratios that are neither the highest nor the lowest.
- Professional/Executive (0.177) and Office Workers (0.118): These groups have the lowest default ratios, suggesting a lower risk of default compared to Sales and Self-employed categories.

**Number of Delinquent Lines**

Analyzing the number of delinquent credit lines provides further insight into risk assessment:

- There is a clear trend showing that as the number of delinquent lines increases, the ratio of defaulted MortDue amount to non-defaulted MortDue amount also rises. This indicates that borrowers with more delinquent lines are at a higher risk of default and potentially larger losses.

**Years at Present Job**

The duration of employment at the current job also affects the default risk:

- Generally, the ratio of defaulted to non-defaulted MortDue decreases as the number of years at the current job increases. This trend suggests that longer job tenure is associated with lower default risk.
- An exception is noted in the 3-5 years range, where there is a slight increase in the default ratio, suggesting a potential area for further investigation.
- Borrowers with over 30 years at their present job exhibit the highest default ratio among long-tenured employees. This anomaly indicates the need for additional analysis, potentially incorporating factors like age and other demographic characteristics, to better understand the underlying causes.

By examining these characteristics—occupational categories, the number of delinquent reports, and job tenure—financial institutions can more effectively identify higher and lower risk credit portfolio assets. This nuanced understanding enables the development of targeted risk mitigation strategies and improves overall portfolio management.

The descriptive dataset analysis with breakdowns by different significant characteristics like Years at present job, profession, etc. are presented in the interactive Looker Data studio report here.

## 5. Key Insights

**Standardized and IRB approach comparison**

- Capital Requirement: The capital requirement under the IRB Approach is significantly higher ($37,3 M) compared to the Standardized Approach ($9,2M). This indicates that the IRB Approach assigns a much higher capital buffer for non-defaulted loans based on its risk assessment.
- Risk Weight (RW): The RW under the IRB Approach is also notably higher (141.44%) compared to the Standardized Approach (34.75%). This suggests that the IRB Approach assigns a higher risk weight to non-defaulted loans, reflecting a higher perceived risk compared to the Standardized Approach.

- Total RWA: The total Risk-Weighted Assets (RWA) under the IRB Approach are substantially higher ($466,8 M) compared to the Standardized Approach ($114,7M). This indicates that the IRB Approach results in a larger capital requirement due to a higher assessment of risk associated with the loan portfolio.

In general standardized approach shows lower capital requirements and lower RWA though IRB aproach is more adjustable and lower capital requirements and RWA might be reached by improving clients portfolio with lower risk of default and implement credit risk mitigation techniques.

**RWA usage and decrease recommendations**
RWA might be improving by improving of data quality and model accuracy, optimize portfolio diversification, enhance credit rating system and applying other measures

**Identification of higher and lower risk credit portfolio assets**

- By occupational categories Sales (0,483) and Self (0,323) categories have the highest defaulted/ non-defaulted MortDue ratios, Other (0,234) and Mgr (0,237) categories have moderate ratios and ProfExe (0,177) and Office (0,118) categories have lowest ratios compared to Sales and Self.

- Portfolio analysis by the number of delinquent lines shows that when number of delinquent lines increases, the ratio of defaulted MortDue amount to non-defaulted MortDue amount rises, indicating a higher risk of default and bigger losses.
- While analyzing by years on the present job the ratio defaulted / non-defaulted MortDue generally decreases as the years at the present job increase, except for a slight increase in the 3-5 years range. There is also highest ratio for the borrowers with more than 30 years at their present job but for this categories more characteristics like age, etc. need to be explored.