SK\_ID\_CURR ID of the loan in our sample TARGET Target variable, this is the future information. Will this loan applicant has payment difficulties? (1: client with payment difficulties: he/she had late payment more than X days, 0: no payment difficulties) CODE\_GENDER Gender of the client FLAG\_OWN\_CAR Flag if the client owns a car FLAG\_OWN\_REALTY Flag if the client owns a house or flat CNT\_CHILDREN Number of children the client has AMT\_INCOME\_TOTAL Income of the client AMT\_CREDIT Credit amount of the loan AMT\_ANNUITY Loan annuity AMT\_GOODS\_PRICE For consumer loans it is the price of the goods for which the loan is given NAME\_TYPE\_SUITE Who was accompanying client when he was applying for the loan NAME\_INCOME\_TYPE Clients income type (businessman, working, maternity leave,…) NAME\_EDUCATION\_TYPE Level of highest education the client achieved NAME\_FAMILY\_STATUS Family status of the client NAME\_HOUSING\_TYPE What is the housing situation of the client (renting, living with parents, ...) DAYS\_BIRTH Client's age in days at the time of application DAYS\_EMPLOYED How many days before the application the person started current employment OCCUPATION\_TYPE What kind of occupation does the client have EXT\_SOURCE\_1 Normalized credit score from an external data source EXT\_SOURCE\_2 Normalized credit score from an external data source EXT\_SOURCE\_3 Normalized credit score from an external data source The “bureau” table stores the other loans of the applicants from the other lenders. “bureau” table: SK\_ID\_CURR ID of loan in our sample - one loan in our sample can have 0,1,2 or more related previous credits in credit bureau SK\_BUREAU\_ID Recoded ID of previous Credit Bureau credit related to our loan (unique coding for each loan application), The IDs of the “other loans” CREDIT\_DAY\_OVERDUE Number of days past due on CB credit at the time of application for related loan in our sample AMT\_CREDIT\_MAX\_OVERDUE Maximal amount overdue on the Credit Bureau credit so far (at application date of loan in our sample) CNT\_CREDIT\_PROLONG How many times was the Credit Bureau credit prolonged AMT\_CREDIT\_SUM Current credit amount for the Credit Bureau credit AMT\_CREDIT\_SUM\_DEBT Current debt on Credit Bureau credit AMT\_CREDIT\_SUM\_LIMIT Current credit limit of credit card reported in Credit Bureau AMT\_CREDIT\_SUM\_OVERDUE Current amount overdue on Credit Bureau credit CREDIT\_TYPE Type of Credit Bureau credit (Car, cash,...) DAYS\_CREDIT\_UPDATE How many days before loan application did last information about the Credit Bureau credit come AMT\_ANNUITY Annuity of the Credit Bureau credit Task 1 : conduct research and analysis to see what factors from the “application” and the “bureau” tables are affecting ● The Credit Scores ● The Payment Difficulty Task 2: Find factors that you can join or make relationship with from 'application' and 'bureau'

### Task 1: Research and Analysis on Factors Affecting Credit Scores and Payment Difficulties

#### A. Factors Affecting Credit Scores

2. \*\*Existing Debt (AMT\_CREDIT\_SUM, AMT\_CREDIT\_SUM\_DEBT)\*\*:

- Total debt amounts can negatively affect credit scores, particularly if they exceed acceptable limits relative to income.

3. \*\*Payment History (CREDIT\_DAY\_OVERDUE, AMT\_CREDIT\_MAX\_OVERDUE)\*\*:

- A record of late payments or significant overdue amounts can significantly lower credit scores.

4. \*\*Loan Amount (AMT\_CREDIT)\*\*:

- Larger loan amounts can increase risk, negatively impacting credit scores if they strain the borrower’s finances.

5. \*\*Annuity Payments (AMT\_ANNUITY)\*\*:

- Higher monthly repayment obligations can be associated with reduced scores if they cause financial strain.

6. \*\*Prolongation of Loans (CNT\_CREDIT\_PROLONG)\*\*:

- Frequent extensions or delays in repayment may indicate financial distress, negatively impacting credit scores.

7. \*\*Credit Types (CREDIT\_TYPE)\*\*:

- Different types of credit (e.g., personal, auto) can have varying effects on credit scores based on historical performance.

8. \*\*Employment Duration (DAYS\_EMPLOYED)\*\*:

- Longer periods of employment generally correlate with higher credit scores, reflecting financial stability.

9. \*\*External Credit Scores (EXT\_SOURCE\_1, EXT\_SOURCE\_2, EXT\_SOURCE\_3)\*\*:

- These scores from external sources provide insights into creditworthiness and can significantly impact overall credit scores.

#### B. Factors Affecting Payment Difficulties

1. \*\*Loan Amount (AMT\_CREDIT)\*\*:

- Larger loans may lead to greater difficulty in repayment if they exceed what borrowers can manage.

2. \*\*Income Level (AMT\_INCOME\_TOTAL)\*\*:

- Lower incomes are generally associated with higher risks of payment difficulties.

5. \*\*Credit Score (EXT\_SOURCE scores)\*\*:

- Lower scores often correlate with a higher likelihood of payment difficulties.

6. \*\*Loan Annuity (AMT\_ANNUITY)\*\*:

- Higher monthly payments can lead to increased risk of missed payments.

7. \*\*Days Overdue on Previous Loans (CREDIT\_DAY\_OVERDUE)\*\*:

- A pattern of overdue payments can indicate future risks of payment difficulties.

8. \*\*Credit History Length (DAYS\_CREDIT\_UPDATE)\*\*:

- A shorter credit history may be associated with higher payment difficulties.

9. \*\*Income Type (NAME\_INCOME\_TYPE)\*\*:

- Certain income types (e.g., freelance work) may be linked to higher risks of payment difficulties.

10. \*\*Family Status (NAME\_FAMILY\_STATUS)\*\*:

- Certain family dynamics may influence financial stability and the likelihood of payment difficulties.

### Task 2: Factors to Join or Create Relationships

1. \*\*SK\_ID\_CURR\*\*:

- This unique identifier serves as the key to join the two tables, linking applications with their corresponding credit bureau records.

2. \*\*Loan Amounts\*\*:

- \*\*AMT\_CREDIT\*\* (from the application table) can be related to \*\*AMT\_CREDIT\_SUM\*\* (from the bureau table) to analyze how total loan amounts influence repayment behaviors.

3. \*\*Payment Behavior\*\*:

- \*\*CREDIT\_DAY\_OVERDUE\*\* (from the bureau table) can be linked with \*\*TARGET\*\* (from the application table) to evaluate how previous overdue payments impact future payment difficulties.

4. \*\*Credit Annuities\*\*:

- By comparing \*\*AMT\_ANNUITY\*\* (from both tables), one can assess how repayment amounts affect payment behavior.

5. \*\*Ownership Status\*\*:

- \*\*FLAG\_OWN\_REALTY\*\* and \*\*FLAG\_OWN\_CAR\*\* from the application table can be analyzed alongside \*\*CREDIT\_TYPE\*\* from the bureau table to explore how ownership affects credit types and repayment.

6. \*\*Employment Stability\*\*:

- Analyzing \*\*DAYS\_EMPLOYED\*\* (from the application table) alongside \*\*DAYS\_CREDIT\_UPDATE\*\* (from the bureau table) can provide insights into how job stability influences credit behavior.

7. \*\*External Credit Scores\*\*:

- \*\*EXT\_SOURCE\_1, EXT\_SOURCE\_2, EXT\_SOURCE\_3\*\* (from the application table) can be correlated with the payment history in the bureau table to evaluate overall creditworthiness.

By examining these factors and their relationships, a deeper understanding of how different variables influence credit scores and payment difficulties can be attained, enabling better risk assessment and lending strategies.

**Task 1: Research and Analysis of Factors Affecting Credit Scores and Payment Difficulties**

**A. Factors Affecting Credit Scores**

1. **Loan Characteristics:**
   * **CREDIT AMOUNT (AMT\_CREDIT):** Larger loans may influence credit utilization rates, impacting credit scores.
   * **ANNUITY (AMT\_ANNUITY):** Regular payments can demonstrate reliability to credit agencies.
   * **PURPOSE (AMT\_GOODS\_PRICE):** Different types of loans may have different risk profiles, affecting credit scores.
2. **Credit History:**
   * **CREDIT OVERDUE (CREDIT\_DAY\_OVERDUE):** High overdue days negatively affect credit scores.
   * **MAX OVERDUE (AMT\_CREDIT\_MAX\_OVERDUE):** Indicates serious issues in past payments.
   * **CREDIT PROLONGS (CNT\_CREDIT\_PROLONG):** Frequent prolongation can signify financial distress, negatively affecting scores.
3. **External Scores:**
   * **EXT\_SOURCE\_1, EXT\_SOURCE\_2, EXT\_SOURCE\_3:** These normalized scores are often strong predictors of creditworthiness and can be directly related to credit scores.

**B. Factors Affecting Payment Difficulties**

1. **Demographics and Personal Factors:**
   * **EDUCATION LEVEL (NAME\_EDUCATION\_TYPE):** Lower education levels might correlate with higher payment difficulties due to potential income issues.
2. **Financial Health:**
   * **INCOME (AMT\_INCOME\_TOTAL):** Lower income may lead to a higher risk of payment difficulties.
3. **Loan Characteristics:**
   * **CREDIT AMOUNT (AMT\_CREDIT):** Higher loans relative to income can increase the likelihood of difficulties.
   * **ANNUITY (AMT\_ANNUITY):** Higher monthly payments might strain finances.
4. **Credit History:**
   * **CURRENT DEBT (AMT\_CREDIT\_SUM\_DEBT):** More debt can lead to higher chances of payment difficulties.
   * **OVERDUE AMOUNT (AMT\_CREDIT\_SUM\_OVERDUE):** Indicates ongoing financial trouble.

**Task 2: Finding Relationships Between 'Application' and 'Bureau' Tables**

1. **Joining on SK\_ID\_CURR:**
   * Both tables share SK\_ID\_CURR, allowing us to merge them based on this key to analyze relationships across both datasets.
2. **Factors to Analyze Together:**
   * **Credit Scores (EXT\_SOURCE\_1, EXT\_SOURCE\_2, EXT\_SOURCE\_3) vs. Bureau Data:** Compare external scores with the overdue data (e.g., CREDIT\_DAY\_OVERDUE, AMT\_CREDIT\_SUM\_OVERDUE).
   * **Income and Credit Overdue:** Analyze how AMT\_INCOME\_TOTAL correlates with CREDIT\_DAY\_OVERDUE and AMT\_CREDIT\_SUM\_DEBT.
   * **Loan Amounts:** Examine how AMT\_CREDIT in the application table relates to AMT\_CREDIT\_SUM and AMT\_CREDIT\_SUM\_DEBT in the bureau table.
   * **Loan Types:** Explore relationships between NAME\_TYPE\_SUITE in the application table and CREDIT\_TYPE in the bureau table to see if certain loan types correlate with higher default risks.
3. **Aggregating Data:**
   * Calculate aggregated statistics (e.g., mean, median) for income and credit scores based on the payment difficulty status to identify trends.
   * Use visualization techniques (like scatter plots or heatmaps) to depict relationships between key variables.

**Conclusion**

This structured analysis allows for a deeper understanding of the factors influencing credit scores and payment difficulties, providing insights that can inform credit risk assessments. Relationships between the two tables will enhance the predictive modeling and risk management strategies for loan approvals.