# **Credit Risk DA Project**

## **Database Connection:**

Download the DBeaver SQL client to connect to the MySQL database:

* <https://dbeaver.io/>

Follow the documentation to set up a connection to the database:

* <https://dbeaver.com/docs/wiki/Create-Connection/>

The database is hosted on AWS, here are the connection details:

* Endpoint: home-credit-default-risk.c7rizeij2t53.ap-southeast-1.rds.amazonaws.com
* Port: 3306
* Database: credit
* Login User: student
* Login Password: student

## **Overview:**

Consider you are asked to review a list of loan applications. The given “credit” database contains data on the loan applicant and their historical loan behavior. There are many columns in the database, you **don’t need to use all the columns**, we will provide a list of useful column descriptions for you.

## **Cautions:**

### **Missing Values:**

* There are columns with missing values. You need to handle them during your analysis. There are multiple ways we can handle missing values: [4 Ways to Replace NULL with a Different Value in MySQL](https://database.guide/4-ways-to-replace-null-with-a-different-value-in-mysql/)

### **Discretization:**

* Discretization means we want to convert numbers into bins, for example, age to age groups or income to income groups. There are mainly 2 reasons for this:
* It is easier to see patterns with a group of values. For example, it is better to say people older than 20 are richer than people younger than 20, instead of saying people aged 20 are richer than people aged 21.
* We want to avoid biased statistics. If we apply group by aggregation directly on a number column like age, the average statistics can be biased. For example, if there is only 1 person aged 59, then the average income of people aged 59 only represents that 1 person in the dataset.

We can do it with the CASE Function in MySQL:

[MySQL CASE Function](https://www.w3schools.com/sql/func_mysql_case.asp)

During the analysis, you can consider converting some factors into groups.

**Task 1: Run SQL via DBeaver**

Follow the documentation to open the “SQL Editor”:

* <https://dbeaver.com/docs/wiki/SQL-Editor/>

Run SQL to examine the number of rows in each table:

|  |  |
| --- | --- |
| **Table** | **Count** |
| application | 307,511 |
| bureau | 1,716,428 |

## 

## **Loan Applications**

The “application” table stores the loan applications. This includes:

* The demographic of the loan applicants
* The loan size or purposes
* The applicant’s credit score
* Is the loan applicant has a payment difficulties with the loan.

|  |  |
| --- | --- |
| 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 |

### **Task 2: What is a Credit Score**

In the “application” table above there are 3 credit score columns. Research online to see what is a credit score and why we need it. (Note that the scores in the database are normalized, which means they are scaled to the 0 to 1 range)

|  |
| --- |
| A **credit score** is a numerical representation of a person's creditworthiness, indicating how likely they are to repay borrowed funds. It is derived from various factors in a person's credit history and helps lenders assess risk when considering a loan application. |

### **Task 3: Understand Credit Amount and Annuity**

What are Credit Amount and Annuity? Fill in your answer below:

|  |  |
| --- | --- |
| Credit Amount | **Credit Amount** refers to the total sum of money that a borrower is authorized to borrow from a lender. It can encompass various types of loans, such as personal loans, mortgages, or credit lines. The credit amount can be influenced by factors like the borrower’s credit score, income, and repayment history. Lenders evaluate these factors to determine how much credit they are willing to extend to the borrower. |
| Annuity | **An annuity** is a financial product that provides a series of payments made at equal intervals. Annuities are commonly used for retirement planning, where individuals can invest a lump sum of money and receive periodic payments over time. |

### **Task 4: Deduce the Loan Duration**

Given the information from Task 4, we should be able to deduce the Loan Duration for each application. Loan duration describes how many periods (months) the applicant will need to pay back their loans.

Paste the SQL and part of the results below:

|  |
| --- |
| **SELECT**  SK\_ID\_CURR,  TARGET,  AMT\_CREDIT,  AMT\_ANNUITY,  **CASE**  **WHEN** AMT\_ANNUITY > 0 **AND** AMT\_CREDIT > 0 **THEN** **ROUND**(AMT\_CREDIT / AMT\_ANNUITY, 2)  **ELSE** **NULL**  **END** **AS** *Loan\_Duration\_Months*  **FROM**  application  **WHERE**  AMT\_CREDIT **IS** **NOT** **NULL**  **AND** AMT\_ANNUITY **IS** **NOT** **NULL**; |

### **Task 5: Are there any factors in the application table affecting the Credit Scores?**

In the “application” table try to explore if there are any columns affecting the credit score. For example, is gender a factor?

**Do the analysis of at least 3 factors for 3 different credit scores**, it is expected to see different results for different credit scores, for example, a factor might affect EXT\_SOURCE\_1 but not EXT\_SOURCE\_3.

Please explain your findings with SQL statements and results:

|  |
| --- |
| 1. **Analyzing Gender**   **SELECT**  **CASE**  **WHEN** CODE\_GENDER = **'F'** **THEN** **'Female'**  **WHEN** CODE\_GENDER = **'M'** **THEN** **'Male'**  **ELSE** **'Unknown'**  **END** **AS** *Gender*,  **COUNT**(\*) **AS** *Total\_Applicants*,  **ROUND**(**AVG**(EXT\_SOURCE\_1), 4) **AS** *Avg\_Ext\_Source\_1*,  **ROUND**(**AVG**(EXT\_SOURCE\_2), 4) **AS** *Avg\_Ext\_Source\_2*,  **ROUND**(**AVG**(EXT\_SOURCE\_3), 4) **AS** *Avg\_Ext\_Source\_3*  **FROM**  application  **WHERE**  CODE\_GENDER **IS** **NOT** **NULL** **AND**  EXT\_SOURCE\_1 **IS** **NOT** **NULL** **AND**  EXT\_SOURCE\_2 **IS** **NOT** **NULL** **AND**  EXT\_SOURCE\_3 **IS** **NOT** **NULL**  **GROUP** **BY**  *Gender*;  **Output:**   **Observations:****Applicant Distribution**  * **Total Applicants**: * **Male**: 34,483 * **Female**: 75,104 * **Unknown**: 2   The data shows a significantly higher number of female applicants compared to male applicants, with females making up approximately 68.5% of the total. This disparity might suggest a higher propensity for females to apply for loans, or it could indicate demographic trends in loan-seeking behavior.   * **Average External Sources** * **Average EXT\_SOURCE\_1**: * **Male**: 0.4118 * **Female**: 0.5523 * **Unknown**: 0.5290   Female applicants have a notably higher average score on EXT\_SOURCE\_1 (0.5523) compared to male applicants (0.4118). This may indicate that female applicants, on average, have better creditworthiness or a more favorable credit history, making them lower risk for lenders.   * **Average EXT\_SOURCE\_2**: * **Male**: 0.5275 * **Female**: 0.5302 * **Unknown**: 0.6589   Both genders have similar scores on EXT\_SOURCE\_2, with females slightly higher than males. The "Unknown" category shows an unusually high score (0.6589), but due to the small sample size (only 2 applicants), it’s not reliable for drawing conclusions.   * **Average EXT\_SOURCE\_3**: * **Male**: 0.4924 * **Female**: 0.4987 * **Unknown**: 0.2188   The average scores for EXT\_SOURCE\_3 are quite close for males and females, indicating that in this particular measure, gender does not have a strong impact. However, the "Unknown" category again has a notably low score, suggesting these cases may represent outliers or unique circumstances not reflected in the larger trends.  **Implications for Lending Practices:**   * **Risk Assessment**: The higher average scores for female applicants in EXT\_SOURCE\_1 and EXT\_SOURCE\_2 suggest that lenders might consider tailoring their risk assessment models. The data indicates that female applicants might present a lower risk on average, which could influence lending strategies and interest rates. * **Focus on Male Applicants**: Given that male applicants have a lower average credit score in EXT\_SOURCE\_1, lenders may want to investigate whether there are underlying reasons for this discrepancy. This could involve targeted financial literacy programs or support to help improve credit scores for male applicants.   **Consideration of Unknowns:**   * The "Unknown" category represents a very small sample size, which limits its usefulness for broader analysis. However, it highlights the importance of ensuring accurate data collection, especially regarding gender identification, to avoid misrepresentation of the applicant pool.   **Potential for Further Analysis:**   * **Demographic Correlations**: Further analysis could explore how other demographic factors (age, income, occupation) intersect with gender in influencing credit scores. This could provide deeper insights into systemic trends. * **Behavioral Trends**: Investigating why women are applying more frequently could reveal behavioral insights that could inform marketing and outreach strategies   **Conclusion:**  Overall, the analysis shows significant differences between male and female applicants in terms of external credit scores.   1. **Analyzing Income Type**   **SELECT**  NAME\_FAMILY\_STATUS **AS** *Family\_Status*,  **COUNT**(\*) **AS** *Total\_Applicants*,  **SUM**(**CASE** **WHEN** EXT\_SOURCE\_1 **IS** **NULL** **THEN** 1 **ELSE** 0 **END**) **AS** *Total\_Null\_Ext\_Source\_1*,  **SUM**(**CASE** **WHEN** EXT\_SOURCE\_2 **IS** **NULL** **THEN** 1 **ELSE** 0 **END**) **AS** *Total\_Null\_Ext\_Source\_2*,  **SUM**(**CASE** **WHEN** EXT\_SOURCE\_3 **IS** **NULL** **THEN** 1 **ELSE** 0 **END**) **AS** *Total\_Null\_Ext\_Source\_3*,  **ROUND**(**AVG**(**COALESCE**(EXT\_SOURCE\_1, 0)), 4) **AS** *Avg\_Ext\_Source\_1*,  **ROUND**(**AVG**(**COALESCE**(EXT\_SOURCE\_2, 0)), 4) **AS** *Avg\_Ext\_Source\_2*,  **ROUND**(**AVG**(**COALESCE**(EXT\_SOURCE\_3, 0)), 4) **AS** *Avg\_Ext\_Source\_3*  **FROM**  application  **WHERE**  NAME\_FAMILY\_STATUS **IS** **NOT** **NULL**  **GROUP** **BY**  *Family\_Status*  **ORDER** **BY**  *Total\_Applicants* **DESC**;  **Output:**   **Observations:**  * **Total Applicants:** * **Population Size**: Larger populations (like "Working") with a number of 158,724 applicants may indicate more stable employment sectors and a broader base for products/services targeting financial literacy and stability. * **Market Segmentation**: Smaller populations (like "Maternity leave") suggest niche markets. Tailored offerings could yield higher engagement from these groups despite their size. * **Resource Allocation**: Understanding the number of applicants helps prioritize which segments require more focus and resources based on their size and needs. * **Null Values for External Scores:**  1. **Total\_Null\_Ext\_Source\_1:** This column shows the count of applicants with null values for EXT\_SOURCE\_1.  * **Insights**: * **Creditworthiness Indicator**: A high count (like in "Working" with 84,455) suggests a significant portion of this population lacks established credit history or financial behavior tracking, indicating potential vulnerabilities in their financial health. * **Risk Assessment**: This group may face challenges in obtaining credit or loans. Identifying these individuals can guide targeted financial education programs to improve their creditworthiness. * **Actionable Steps**: Programs aimed at helping applicants build credit (such as secured credit cards or financial coaching) could be beneficial.  1. **Total\_Null\_Ext\_Source\_2:** This metric shows the number of applicants with null values for EXT\_SOURCE\_2.  * **Insights**: * **Financial Insights Gap**: The presence of null values (e.g., 315 in "Working") indicates missing data that could be crucial for assessing financial behavior. This can affect risk profiling. * **Targeted Data Collection**: Strategies to improve data completeness, such as follow-ups or incentives for applicants to provide comprehensive financial information, could enhance future analyses. * **Correlation with Financial Literacy**: Those with null values may lack awareness of the importance of these external scores, highlighting a need for educational initiatives.  1. **Total\_Null\_Ext\_Source\_3:** This column indicates the number of applicants with null values for EXT\_SOURCE\_3.  * **Insights**: * **Potential Blind Spots**: High null counts (e.g., 31,780 in "Working") mean that crucial aspects of financial behavior are untracked. This can skew analysis and hinder risk assessments. * **Data Quality Improvement**: Addressing data gaps can lead to more accurate financial assessments. Initiatives to encourage data provision can yield a more comprehensive view of applicants. * **Impact on Product Offerings**: A lack of EXT\_SOURCE\_3 data might indicate that products tailored to these individuals may not be effectively aligned with their financial profiles. * **Detailed Insights by Income Type**  1. **Working:** High null count for EXT\_SOURCE\_1 (84,455)  * **Average EXT\_SOURCE\_1**: 0.2185 * **Average EXT\_SOURCE\_2**: 0.5009 * **Average EXT\_SOURCE\_3**: 0.3987 * **Insights**: * **Financial Stability**: The average EXT\_SOURCE\_1 score indicates relatively low financial stability among the working population. This could point to issues such as high debt-to-income ratios or unstable income sources. * **Opportunities**: Given their large size, this segment presents an opportunity for financial products designed to enhance financial literacy and budgeting skills. Consider workshops, online resources, or mobile apps that help them manage their finances more effectively. * **Targeted Outreach**: Given the high number of null values, outreach initiatives aimed at those with incomplete credit histories may help these individuals build better financial profiles.  1. **Commercial Associate:** Moderate counts for EXT\_SOURCE\_2 (141) and EXT\_SOURCE\_3 (14,871)  * **Average EXT\_SOURCE\_1**: 0.251 * **Average EXT\_SOURCE\_2**: 0.5411 * **Average EXT\_SOURCE\_3**: 0.3962 * **Insights**: * **Moderate Financial Health**: This group displays better average scores than the "Working" category, suggesting that they have slightly better financial management or resources. * **Targeted Financial Products**: Financial products that encourage savings or investment could be appealing to this demographic. They may benefit from investment options that can leverage their moderate financial standing. * **Risk Awareness**: Continuous monitoring of financial behaviors can help identify those at risk of falling into financial distress, enabling proactive support.  1. **Pensioner:** High null count for EXT\_SOURCE\_1 (42,512)  * **Average EXT\_SOURCE\_1**: 0.1609 * **Average EXT\_SOURCE\_2**: 0.505 * **Average EXT\_SOURCE\_3**: 0.4473 * **Insights**: * **Potential Vulnerabilities**: The pensioner category has the lowest average EXT\_SOURCE\_1, indicating potential vulnerabilities. They might rely solely on fixed incomes, making them more susceptible to economic changes. * **Financial Security Products**: There is an opportunity to offer products focused on financial security, such as annuities or insurance products that provide peace of mind. Education on managing fixed incomes can also be beneficial. * **Support Services**: Programs that help pensioners manage their finances or navigate healthcare costs can significantly impact their quality of life.  1. **State Servant:** Moderate null count for EXT\_SOURCE\_2 (48)  * **Average EXT\_SOURCE\_1**: 0.2656 * **Average EXT\_SOURCE\_2**: 0.5334 * **Average EXT\_SOURCE\_3**: 0.4378 * **Insights**: * **Stable Income**: State servants generally have more stable income compared to other groups, which may contribute to a higher average in EXT\_SOURCE\_2. * **Retirement Planning**: As this group approaches retirement, tailored financial planning services focusing on retirement savings, pension maximization, and investment could be beneficial. * **Job Security**: Emphasizing job security in communications can enhance trust and engagement with this demographic, providing opportunities for loyalty programs.  1. **Unemployed:** Significant count for EXT\_SOURCE\_1 (15)  * **Average EXT\_SOURCE\_1**: 0.1326 * **Average EXT\_SOURCE\_2**: 0.4593 * **Average EXT\_SOURCE\_3**: 0.1054 * **Insights**: * **High Financial Risk**: This group has the lowest average scores across all external sources, indicating a significant risk of financial instability. * **Intervention Programs**: There's a pressing need for intervention programs focused on job placement, vocational training, and financial counseling. Supporting unemployed individuals in gaining skills can help them re-enter the workforce. * **Social Safety Nets**: Engaging with local community programs or government initiatives to offer support services could improve outcomes for this demographic.  1. **Student:** Low counts for EXT\_SOURCE\_1 (9)  * **Average EXT\_SOURCE\_1**: 0.2103 * **Average EXT\_SOURCE\_2**: 0.500 * **Average EXT\_SOURCE\_3**: 0.4175 * **Insights**: * **Emerging Needs**: Students are often in transitional phases regarding financial independence. They may have unique needs related to education financing, budgeting, and future employment. * **Financial Education**: Providing financial education programs focused on managing student loans and budgeting could be beneficial. Tools tailored to young adults that teach financial literacy can foster better financial habits early on. * **Potential Long-Term Engagement**: Building relationships with students now may lead to lifelong customers. Introducing student-oriented banking products can help cultivate brand loyalty.  1. **Businessman:** Moderate count for EXT\_SOURCE\_1 (6)  * **Average EXT\_SOURCE\_1**: 0.522 * **Average EXT\_SOURCE\_2**: 0.6664 * **Average EXT\_SOURCE\_3**: 0.5629 * **Insights**: * **Higher Financial Capacity**: This group has relatively higher external scores, suggesting better financial health and resource availability. * **Growth Potential**: Businessmen often seek products that facilitate growth. Offering tailored loans, investment advice, and networking opportunities can attract this demographic. * **Entrepreneurial Support**: Providing resources for entrepreneurship, such as workshops and mentorship programs, can enhance engagement and loyalty.  1. **Maternity Leave**  * **Average EXT\_SOURCE\_1**: 0.5329 * **Average EXT\_SOURCE\_2**: 0.5121 * **Average EXT\_SOURCE\_3**: 0.0 (null count for EXT\_SOURCE\_3 is significant) * **Insights**: * **Limited Representation**: With only 5 applicants, this category may represent a neglected segment. The lack of data in EXT\_SOURCE\_3 indicates a gap in understanding their financial situation. * **Targeted Support**: Developing programs that cater specifically to those on maternity leave can fill this gap. Financial planning tools and temporary financial assistance products could be beneficial. * **Future Engagement**: Engaging with this demographic during maternity can lead to long-term customer loyalty, as they transition back into the workforce.   **Conclusions:**   * **Risk Profiles by Income Type:** * **Vulnerability Insights:** Different income types exhibit varying levels of financial stability and risk. For instance, the "Working" group, despite its large size, has a high number of null values in external scores, indicating potential creditworthiness issues. Conversely, pensioners show low average scores in EXT\_SOURCE\_1, reflecting their reliance on fixed incomes and susceptibility to economic fluctuations. * **Implications for Lending:** * **Customized Lending Strategies:** The varying average external scores across income types suggest the need for customized lending strategies. For example, the "Unemployed" group presents a high financial risk, indicating that traditional lending approaches may not be suitable. Instead, institutions might consider offering microloans or financial education alongside credit products to support this demographic. * **Potential Areas for Further Research:** * **Data Completeness and Quality:** Investigating the causes behind the high number of null values in external scores can provide insights into the data collection processes. Understanding these gaps will inform strategies to enhance data completeness and accuracy, which are crucial for robust financial analysis and decision-making.  1. **Analyzing Family Status**   **SELECT**  NAME\_FAMILY\_STATUS **AS** *Family\_Status*,  **COUNT**(\*) **AS** *Total\_Applicants*,  **SUM**(**CASE** **WHEN** EXT\_SOURCE\_1 **IS** **NULL** **THEN** 1 **ELSE** 0 **END**) **AS** *Total\_Null\_Ext\_Source\_1*,  **SUM**(**CASE** **WHEN** EXT\_SOURCE\_2 **IS** **NULL** **THEN** 1 **ELSE** 0 **END**) **AS** *Total\_Null\_Ext\_Source\_2*,  **SUM**(**CASE** **WHEN** EXT\_SOURCE\_3 **IS** **NULL** **THEN** 1 **ELSE** 0 **END**) **AS** *Total\_Null\_Ext\_Source\_3*,  **ROUND**(**AVG**(**COALESCE**(EXT\_SOURCE\_1, 0)), 4) **AS** *Avg\_Ext\_Source\_1*,  **ROUND**(**AVG**(**COALESCE**(EXT\_SOURCE\_2, 0)), 4) **AS** *Avg\_Ext\_Source\_2*,  **ROUND**(**AVG**(**COALESCE**(EXT\_SOURCE\_3, 0)), 4) **AS** *Avg\_Ext\_Source\_3*  **FROM**  application  **WHERE**  NAME\_FAMILY\_STATUS **IS** **NOT** **NULL**  **GROUP** **BY**  *Family\_Status*  **ORDER** **BY**  *Total\_Applicants* **DESC**;  **Output:**    **Observations:**   * **Total Applicants:** * Married individuals represent the largest group, with 196,432 applicants, indicating that this demographic likely has the most stable financial circumstances and is potentially more engaged in financial services. * Single/not married applicants are significantly fewer, at 45,444, while the "Unknown" category has only 2 applicants, indicating possible data quality issues. * **Null Values for External Scores:** * The total of null values in **EXT\_SOURCE\_1** are notably high, especially for the **Married** category (**110,244** nulls). This may indicate that a substantial portion of married applicants lacks sufficient credit history, which can complicate risk assessments. * Widows and Separated individuals also exhibit substantial null counts, indicating that these groups may face challenges in obtaining credit or loans due to missing data. * **Detailed Insights by Family Status:**  1. **Married**: Total Applicants (196,432)  * **Avg EXT\_SOURCE\_1**: 0.2245 * **Avg EXT\_SOURCE\_2**: 0.5193 * **Avg EXT\_SOURCE\_3**: 0.4194   **Insights**: Despite the largest applicant pool, the average score for **EXT\_SOURCE\_1** is low, indicating many may lack a robust credit history. The moderate scores in **EXT\_SOURCE\_2** and **EXT\_SOURCE\_3** suggest that while this group is perceived as low to moderate risk. Programs aimed at financial literacy and building credit could be beneficial.   1. **Single / Not Married**: Total Applicants (45,444)  * **Avg EXT\_SOURCE\_1**: 0.2056 * **Avg EXT\_SOURCE\_2**: 0.4959 * **Avg EXT\_SOURCE\_3**: 0.3706   **Insights**: This group’s lower average scores reflect potential financial instability. This may reflect the higher risk associated with individuals without a partner to share financial responsibilities, which might lead to perceived instability. Financial products tailored for budgeting and savings education could help improve their financial situation.   1. **Civil Marriage**: Total Applicants (29,775)  * **Avg EXT\_SOURCE\_1**: 0.2106 * **Avg EXT\_SOURCE\_2**: 0.4964 * **Avg EXT\_SOURCE\_3**: 0.3827   **Insights**: Similar to the single group, the civil marriage group shows modest financial health. Targeted outreach for financial products aimed at relationship-based budgeting could yield positive outcomes.   1. **Separated**: Total Applicants (19,770)  * **Avg EXT\_SOURCE\_1**: 0.2412 * **Avg EXT\_SOURCE\_2**: 0.5231 * **Avg EXT\_SOURCE\_3**: 0.419   **Insights**: This group exhibits relatively better scores compared to others, suggesting that they may have resources or stability post-separation. However, the high null count in **EXT\_SOURCE\_1** indicates that financial institutions should provide educational resources aimed at helping this demographic manage their credit and finances during transitions.   1. **Widow**: Total Applicants (16,088)  * **Avg EXT\_SOURCE\_1**: 0.178 * **Avg EXT\_SOURCE\_2**: 0.5086 * **Avg EXT\_SOURCE\_3**: 0.4379   **Insights**: The lowest average for EXT\_SOURCE\_1 highlights vulnerabilities for this demographic. However, considering there is a slight improvement throughout EXT\_SOURCE\_2 and EXT\_SOURCE\_3 shows that there is potential for widows to improve their financial circumstances. Support services, such as financial counseling and retirement planning, are crucial.   1. **Unknown**: Total Applicants (2)  * **Avg EXT\_SOURCE\_1**: 0.3143 * **Avg EXT\_SOURCE\_2**: 0.6729 * **Avg EXT\_SOURCE\_3**: 0.3353   **Insights**: This group is too small for significant analysis but indicates potential issues in data collection.  **Conclusions:**   * **Differential Risk Assessment**: The findings suggest that lenders should consider family status when assessing creditworthiness. For instance, while married individuals may present a lower risk, significant null values indicate a lack of credit history that could influence lending decisions. * **Targeted Financial Products**: Tailoring financial products to specific family statuses could enhance financial inclusion. For example, products aimed at individuals who are single or widowed could focus on financial education and building credit. * **Data Quality Enhancement**: The high number of null values highlights the need for improved data collection and validation processes. Better data quality will enable lenders to make more informed decisions and reduce risk. |

### **Task 6: Are there any factors in the application table affecting the Credit Amount?**

Who is going to lend more money than others? In this task, we want to see are there any factors affecting the credit amount. **Do the analysis of at least 3 factors**

Please explain your findings with SQL statements and results:

|  |
| --- |
| 1. **Income Level**   **SELECT**  **CASE**  **WHEN** AMT\_INCOME\_TOTAL <= 25000 **THEN** **'0 - 25,000'**  **WHEN** AMT\_INCOME\_TOTAL **BETWEEN** 25001 **AND** 50000 **THEN** **'25,001 - 50,000'**  **WHEN** AMT\_INCOME\_TOTAL **BETWEEN** 50001 **AND** 75000 **THEN** **'50,001 - 75,000'**  **WHEN** AMT\_INCOME\_TOTAL **BETWEEN** 75001 **AND** 100000 **THEN** **'75,001 - 100,000'**  **ELSE** **'100,001 and above'**  **END** **AS** *Income\_Level*,  **AVG**(AMT\_CREDIT) **AS** *Average\_Amount\_Credit*  **FROM**  application  **GROUP** **BY**  *Income\_Level*  **ORDER** **BY**  **CASE**  **WHEN** Income\_Level = **'0 - 25,000'** **THEN** 1  **WHEN** Income\_Level = **'25,001 - 50,000'** **THEN** 2  **WHEN** Income\_Level = **'50,001 - 75,000'** **THEN** 3  **WHEN** Income\_Level = **'75,001 - 100,000'** **THEN** 4  **ELSE** 5  **END**;  **Output:**    **Observation:**   * **25,001 - 50,000**:   + **Average Credit Amount**: **294,668.66**   + **Interpretation**: Borrowers in this income range receive a relatively moderate amount of credit. This suggests that while they have some financial stability, their lower income may lead lenders to offer cautious loan amounts. * **50,001 - 75,000**:   + **Average Credit Amount**: **343,394.31**   + **Interpretation**: As income increases, the average credit amount offered also rises. This indicates that lenders perceive these borrowers as more capable of managing larger loans, likely due to a higher disposable income. * **75,001 - 100,000**:   + **Average Credit Amount**: **419,047.66**   + **Interpretation**: This income bracket shows a significant increase in the average credit amount. It suggests that borrowers in this range are seen as more financially secure, leading to greater trust from lenders to extend larger loans. * **100,001 and above**:   + **Average Credit Amount**: **654,362.93**   + **Interpretation**: Individuals earning above 100,000 have the highest average credit amount. This substantial increase indicates that lenders view these borrowers as the least risky, willing to extend much larger loans based on their financial capacity.  **Overall Trend Analysis:**  * **Positive Correlation**: The results show a clear positive correlation between income and average credit amount. As borrowers' income increases, lenders are more likely to offer larger credit amounts. This trend reflects typical lending behavior, where higher income indicates a greater ability to repay loans. * **Risk Perception**: Lenders likely assess risk based on income, with higher earners considered lower risk. This leads to more favorable loan terms and larger amounts for those in higher income brackets.  **Implications for Borrowers and Lenders:**  * **Borrowers**: Individuals seeking loans should be aware that their income level significantly impacts the amount they can borrow. Higher income may not only lead to larger loans but also better interest rates. * **Lenders**: Financial institutions can use this data to tailor their lending strategies. They might focus on higher-income segments for larger loan products while being more conservative with lower-income borrowers.  1. **Ownership Status**  * **Car Ownership**:   **SELECT**  FLAG\_OWN\_CAR **AS** *Car\_Own*,  **AVG**(AMT\_CREDIT) **AS** *Average\_Amount\_Credit*  **FROM**  application  **GROUP** **BY**  *Car\_Own*;  **Output:**    **Observation:**   * **No Car Ownership (N)**:   + **Average Credit Amount**: **565,442.54**   + **Interpretation**: Borrowers who do not own a car receive a substantial average credit amount. This might suggest that lenders are willing to extend significant loans to these individuals despite the absence of a car, indicating other factors (such as income or credit score) may be compensating for this lack of asset. * **Car Ownership (Y)**:   + **Average Credit Amount**: **664,186.01**   + **Interpretation**: Car owners have an even higher average credit amount. This reflects the perception that owning a car signifies financial stability and responsibility, leading lenders to offer more substantial loans to these borrowers.  **Comparative Analysis:**  * **Difference in Average Credit**: The average credit amount for car owners is **approximately 98,743.47** more than that for non-car owners. This significant difference highlights a trend where car ownership is associated with a higher perceived creditworthiness.  **Implications for Lenders and Borrowers:**  * **For Lenders**:   + **Risk Assessment**: Lenders may view car ownership as a positive indicator of a borrower’s financial reliability. Owning a car can signal that the borrower has made a substantial financial commitment and is likely to have a stable income.   + **Targeting Strategies**: Financial institutions might tailor their marketing strategies to attract car owners, offering them competitive loan products based on their perceived lower risk. * **For Borrowers**:   + **Understanding Loan Offers**: Individuals looking to secure loans may benefit from understanding how their assets, such as car ownership, influence their borrowing capacity. Those without cars might want to emphasize other financial strengths, like stable income or good credit scores, to enhance their creditworthiness in the eyes of lenders. * **Realty Ownership**:   **SELECT**  FLAG\_OWN\_REALTY,  **AVG**(AMT\_CREDIT) **AS** *avg\_amt\_credit*  **FROM**  application  **GROUP** **BY**  FLAG\_OWN\_REALTY;  **Output:**    **Observation:**   * **Realty Ownership (Y)**:   + **Average Credit Amount**: **588,522.52**   + **Interpretation**: Borrowers who own real estate receive an average credit amount of 588,522.52. This indicates that while these borrowers are still seen as fairly creditworthy, the average loan amount is lower compared to those without realty ownership. * **No Realty Ownership (N)**:   + **Average Credit Amount**: **622,810.94**   + **Interpretation**: Interestingly, borrowers without real estate ownership have a higher average credit amount. This suggests that lenders may perceive these individuals as having other financial strengths or lower risk factors that justify larger loan amounts.  **Comparative Analysis:**  * **Difference in Average Credit**: The average credit amount for non-realty owners is approximately **34,288.42** more than that for realty owners. This unexpected finding indicates that owning real estate does not necessarily correlate with higher credit offers.  **Implications for Lenders and Borrowers:**  * **For Lenders**:   + **Risk Assessment**: The data suggests that lenders may have differing perceptions of risk based on real estate ownership. They might view non-realty owners as potentially more financially fluid or having access to other assets or income sources that allow for larger loans.   + **Tailored Lending**: This insight could prompt lenders to reconsider how they evaluate realty ownership in their risk assessments and lending strategies. * **For Borrowers**:   + **Understanding Loan Offers**: Borrowers who own real estate might find it beneficial to showcase other financial strengths (like income or credit history) when applying for loans, as ownership alone doesn’t guarantee higher credit amounts.   + **Market Strategies**: Those without real estate ownership may feel encouraged to apply for larger loans, knowing they might receive favorable terms despite not having realty assets.  1. **Family Status**   **SELECT**  NAME\_FAMILY\_STATUS,  **AVG**(AMT\_CREDIT) **AS** *avg\_amt\_credit*  **FROM**  application  **GROUP** **BY**  NAME\_FAMILY\_STATUS;  **Output:**    **Observation:**   * **Single / Not Married**:   + **Average Credit Amount**: **505,350.18**   + **Interpretation**: Single individuals receive the lowest average credit amount among the categories analyzed. This may suggest that lenders perceive single borrowers as potentially higher risk, possibly due to the lack of shared financial responsibilities or stability that comes with partnerships. * **Married**:   + **Average Credit Amount**: **642,999.79**   + **Interpretation**: Married individuals receive a significantly higher average credit amount. Lenders may view married couples as more stable and capable of managing larger debts, likely due to combined incomes and shared financial responsibilities. * **Civil Marriage**:   + **Average Credit Amount**: **541,573.46**   + **Interpretation**: Borrowers in civil marriages also receive a healthy average credit amount, although it is lower than that for married couples. This may indicate that while lenders see these relationships as stable, they may not perceive them as strongly as traditional marriages. * **Widow**:   + **Average Credit Amount**: **490,695.91**   + **Interpretation**: Widowed individuals receive an average credit amount similar to singles, suggesting that lenders might consider them in a higher risk category, possibly due to the loss of a financial partner. * **Separated**:   + **Average Credit Amount**: **552,113.82**   + **Interpretation**: Those who are separated receive a moderate credit amount. This reflects a somewhat higher risk perception compared to married individuals, but potentially more stability than single or widowed borrowers. * **Unknown**:   + **Average Credit Amount**: **630,000.00**   + **Interpretation**: The average credit amount for borrowers with unknown family status is notably high. This could indicate that these borrowers may possess other favorable attributes (like income or credit score) that lead lenders to extend larger loans despite the lack of clarity on their family situation.  **Comparative Analysis:**  * **Highest Average Credit**: Married individuals have the highest average credit amount, indicating that lenders favor stable family structures. * **Lowest Average Credit**: Single individuals and widows receive the lowest amounts, suggesting higher perceived risk associated with these statuses.  **Implications for Lenders and Borrowers:**  * **For Lenders**:   + **Risk Evaluation**: These findings underscore the importance of family status in assessing risk. Lenders might consider expanding their evaluation criteria to include other financial indicators alongside family status.   + **Tailored Loan Products**: Lenders may develop tailored products for different family statuses, addressing the unique financial needs and risks associated with each group. * **For Borrowers**:   + **Understanding Loan Offers**: Individuals should recognize how their family status may impact their loan eligibility and amounts. This insight can help them prepare better financial profiles or seek additional documentation that showcases their financial stability. |

### 

### **Task 7: Are there any factors in the application table affecting the Payment Difficulties?**

In the database, the **TARGET** column describes will there be a payment difficulty for a loan. We want to see if there are any factors in the application table that can be used to predict this future information. **Do the analysis of at least 3 factors.**

Please explain your findings with SQL statements and results:

|  |
| --- |
| 1. **Home Ownership**   **SELECT**  FLAG\_OWN\_REALTY **AS** *Home\_Ownership*,  **AVG**(TARGET) \* 100 **AS** *Average\_Payment\_Difficulty\_Percentage*  **FROM**  application  **GROUP** **BY**  FLAG\_OWN\_REALTY;  **Output:**    **Observation:**   * **Homeowners (FLAG\_OWN\_REALTY = 'Y')**:   + **Average Payment Difficulty Rate**: **0.0796** (or 7.96%)   + **Interpretation**: Approximately 8% of homeowners experience payment difficulties. This relatively low rate suggests that owning real estate may provide a degree of financial stability, making it easier for these borrowers to manage their loan repayments. * **Non-Homeowners (FLAG\_OWN\_REALTY = 'N')**:   + **Average Payment Difficulty Rate**: **0.0832** (or 8.32%)   + **Interpretation**: About 8.32% of non-homeowners face payment difficulties, which is slightly higher than the rate for homeowners. This indicates a marginally increased risk of payment difficulties for those without real estate.  **Comparative Analysis:**  * The difference in payment difficulty rates between homeowners and non-homeowners is relatively small (0.0036 or 0.36%). However, homeowners have a lower average payment difficulty rate, implying that they are somewhat less likely to face repayment challenges.  **Implications:**  * **Financial Stability**: The findings suggest that owning property may confer some financial advantages that help borrowers manage their debts more effectively. This could be due to the potential for increased net worth, access to equity, or greater perceived stability by lenders. * **Lender Insights**: For lenders, this information is valuable for risk assessment. They might consider including real estate ownership as a factor when evaluating potential borrowers. Homeowners might be viewed as less risky compared to non-homeowners, which could influence lending terms and conditions.  **Conclusion:** Overall, the analysis indicates that while both homeowners and non-homeowners face payment difficulties, homeowners tend to do so at a slightly lower rate. This suggests that home ownership can serve as a buffer against repayment issues, aligning with common perceptions of financial stability associated with owning real estate.   1. **Number of Children**   **SELECT**  CNT\_CHILDREN **AS** *Number\_of\_Children*,  **AVG**(TARGET) \* 100 **AS** *Average\_Payment\_Difficulty\_Percentage*  **FROM**  application  **GROUP** **BY**  CNT\_CHILDREN  **ORDER** **BY**  *Number\_of\_Children*;  **Output:**    **Observation:**  The average payment difficulty rates are as follows:   * **0 Children**: **0.0771** (or 7.71%) * **1 Child**: **0.0892** (or 8.92%) * **2 Children**: **0.0872** (or 8.72%) * **3 Children**: **0.0963** (or 9.63%) * **4 Children**: **0.1282** (or 12.82%) * **5 Children**: **0.0833** (or 8.33%) * **6 Children**: **0.2857** (or 28.57%) * **7 Children**: **0.0000** (or 0%) * **8 Children**: **0.0000** (or 0%) * **9 Children**: **1.0000** (or 100%) * **10 Children**: **0.0000** (or 0%) * **11 Children**: **1.0000** (or 100%) * **12 Children**: **0.0000** (or 0%) * **14 Children**: **0.0000** (or 0%) * **19 Children**: **0.0000** (or 0%)  **Analysis by Number of Children:**  * **No Children (0)**: The payment difficulty rate is the lowest at **7.71%**, indicating that individuals without dependents are generally less likely to face repayment issues. * **One Child (1)**: The rate increases to **8.92%**, suggesting a slight increase in financial burden with the addition of a child. * **Two Children (2)**: The rate decreases slightly to **8.72%**, indicating that having two children may not significantly impact repayment difficulties compared to one child. * **Three Children (3)**: The rate rises to **9.63%**, showing that as the number of children increases, so does the payment difficulty, indicating increased financial strain. * **Four Children (4)**: The average payment difficulty further increases to **12.82%**, reinforcing the idea that larger family sizes correlate with greater repayment challenges. * **Five Children (5)**: Interestingly, the rate drops to **8.33%**, which may suggest a smaller sample size or other mitigating factors for these borrowers. * **Six Children (6)**: The rate spikes dramatically to **28.57%**, suggesting that individuals with six children face significant financial stress, making them much more likely to encounter payment difficulties.  **Extreme Cases:**  * **Seven Children (7)**: The payment difficulty rate is **0%**, which may indicate a specific circumstance where all these borrowers are managing their finances well, or it may also reflect a smaller or unique sample. * **Nine (9) and Eleven Children (11)**: Both categories have a **100%** payment difficulty rate, indicating that every borrower in this group has experienced payment issues. This suggests that families with nine or more children face extreme financial challenges, making repayment virtually impossible for those particular cases. * **Other Cases (8, 10, 12, 14, 19)**: All have a **0%** rate, indicating successful management of repayment, possibly due to unique circumstances in these specific groups or very small sample sizes.  **Overall Trends and Implications:**  * **Increasing Family Size**: Generally, as the number of children increases, so does the payment difficulty rate, suggesting that financial obligations rise significantly with each additional dependent. * **Support for Financial Planning**: These findings can be crucial for lenders. They may want to consider family size as a factor in risk assessment when evaluating loan applications. Larger families may require different loan products or terms to accommodate their unique financial situations. * **Policy Implications**: This data can also inform social services or financial advisory programs aimed at families, helping to provide targeted support for larger households that might struggle more with debt repayment.  **Conclusion:** The analysis clearly indicates that family size, measured by the number of children, has a significant impact on the likelihood of facing payment difficulties. While smaller families may experience lower risks, larger families—especially those with extreme numbers of dependents—face considerable financial strain. |

## **Previous/Other Loan Applications:**

In the previous section, we explored if the demographic data related to payment difficulties, this section we want to see if **historical loan behavior** affecting the payment difficulties.

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 7: Is the number of other loans affecting the payment difficulties?**

We want to see if loan applicants have other historical loans affecting their payment abilities. Hints:

* You will need to count the number of loans for each SK\_ID\_CURR in the “bureau” table.
* Transform the counts into count groups (Discretization).
* Compute the relation between average other loan count to the TARGET

Paste the SQL and part of the results below:

|  |
| --- |
| **SELECT**  **CASE**  **WHEN** *Total\_Loan* = 0 **THEN** **'0 Loans'**  **WHEN** *Total\_Loan* **BETWEEN** 1 **AND** 2 **THEN** **'1-2 Loans'**  **WHEN** *Total\_Loan* **BETWEEN** 3 **AND** 4 **THEN** **'3-4 Loans'**  **WHEN** *Total\_Loan* **BETWEEN** 5 **AND** 6 **THEN** **'5-6 Loans'**  **ELSE** **'7+ Loans'**  **END** **AS** *Loan\_Group*,  **AVG**(*a*.TARGET) \* 100 **AS** *Average\_Payment\_Difficulty\_Percent*,  **COUNT**(*a*.SK\_ID\_CURR) **AS** *Number\_of\_Applicants*  **FROM** (  **SELECT**  SK\_ID\_CURR,  **COUNT**(\*) **AS** *Total\_Loan*  **FROM**  bureau  **GROUP** **BY**  SK\_ID\_CURR  ) **AS** *Total\_Loans*  **JOIN**  application *a* **ON** *Total\_Loans*.SK\_ID\_CURR = *a*.SK\_ID\_CURR  **GROUP** **BY**  *Loan\_Group*  **ORDER** **BY**  *Loan\_Group*;  **Output:**   **Observation:**  * **1-2 Loans:** * **Average Payment Difficulty Rate:** **0.0820 (or 8.20%)** * **Number of Applicants:** **71,707** * **Insights**: This group exhibits the highest average payment difficulty rate among all loan groups. The 8.20% rate indicates that a significant portion of borrowers with only one or two loans encounter challenges in meeting their repayment obligations. * **Possible Factors**:   + - **Inexperience with Debt**: Borrowers with only a couple of loans may be less experienced in managing their debt load. They may not fully understand how to budget effectively, leading to higher risks of default.     - **Financial Strain**: Those in this category might face unexpected financial challenges, such as job loss or medical emergencies, without the cushioning that additional loans or diverse financial products might provide.     - They may also be attempting to take on credit obligations before fully understanding the implications of additional debt. * **3-4 Loans:** * **Average Payment Difficulty Rate**: **0.0738 (or 7.38%)** * **Number of Applicants**: **61,898** * **Analysis**: A noticeable decline in the payment difficulty rate suggests an improvement in financial management among borrowers. This could indicate a growing familiarity with loan management and budgeting. * **Possible Factors**:   + - **Experience and Learning**: Individuals in this bracket may have gained experience in handling their debt, leading to better financial planning and management skills.     - **Financial Stability**: They may also benefit from having a more stable income or better budgeting practices that allow them to meet their obligations more comfortably.     - This group may have access to more financial resources or support systems compared to those with fewer loans.  **5-6 Loans**:  * **Number of Applicants**: **45,937** * **Analysis**: The payment difficulty rate remains stable, indicating that borrowers can manage their financial responsibilities well even with a moderate to high number of loans. * **Possible Factors**:   + - **Debt Management Strategies**: These borrowers may have developed effective debt management strategies, such as consolidating loans or prioritizing higher-interest debts.     - **Greater Financial Acumen**: There may be a stronger financial acumen, enabling them to navigate potential pitfalls of accumulating debt.     - Stability in income sources might also contribute to their ability to manage multiple financial obligations without facing difficulties.  **7+ Loans**:  * **Average Payment Difficulty Rate**: **0.0780 (or 7.80%)** * **Number of Applicants**: **83,949** * **Analysis**: A slight increase in payment difficulty indicates that while many borrowers can manage multiple loans, there’s a threshold where financial stress begins to surface. * **Possible Factors**:   + - **Income Constraints**: Accumulating multiple loans can lead to increased monthly obligations, which may exceed borrowers' financial capacity, especially if their income doesn’t increase correspondingly.     - **Over-Leverage Risk**: Borrowers with seven or more loans may be approaching or exceeding their financial limits, which can lead to increased stress and higher likelihood of default.     - This group may also face external economic pressures (e.g., job loss, medical emergencies) that can disproportionately affect those with numerous financial obligations.  **Overall Trends:**  * **U-Shaped Pattern**: The findings indicate a U-shaped relationship regarding payment difficulties. Initially, borrowers with a limited number of loans (1-2) exhibit higher risk levels, while those with a moderate number of loans (3-4, 5-6) experience decreased difficulties. However, when borrowers exceed six loans, the risks begin to escalate again, suggesting that excessive debt may contribute to repayment challenges.  **Implications for Borrowers:**  * **Financial Education**: Borrowers with 1-2 loans might benefit from financial education and counseling to help them understand the implications of credit and how to manage multiple obligations effectively. * **Support Services**: For those in the higher loan brackets, support services like debt management programs could assist them in navigating their financial landscape more effectively.  **Implications for Lenders:**  * **Risk Assessment Models**: Lenders should consider the number of loans as a crucial variable in their risk assessment algorithms. Recognizing that borrowers with 1-2 loans may present higher risk can significantly inform lending decisions. * **Tailored Loan Offerings**: Lenders should consider developing differentiated loan products that specifically address the needs of borrowers based on their number of existing loans. By providing flexible terms and additional support for those in higher loan brackets, lenders can effectively mitigate the risks of default. * **Monitoring Borrower Behavior**: Continuous monitoring of borrowers' financial behavior, particularly for those nearing the threshold of excessive debt, enables lenders to intervene proactively before payment difficulties occur.  **Conclusion:** The detailed analysis indicates that the relationship between the number of historical loans and payment difficulties is complex and nuanced. While a moderate number of loans can suggest effective financial management, there is a critical threshold beyond which the risk of repayment challenges increases significantly. Lenders can utilize these insights to enhance their strategies and provide better support to borrowers, ultimately promoting improved financial health for both parties. |

## **Task 8: FreeStyle**

Now, conduct your own research and analysis to see what factors from the “application” and the “bureau” tables are affecting

* The Credit Scores
* The Payment Difficulty

**Exploration 1: Research and Analysis on Factors Affecting Credit Scores and Payment Difficulties**

1. **Factors Affecting Credit Scores**

* Demographic Information:

|  |
| --- |
| 1. **AGE (DAYS\_BIRTH):** Older applicants may have more credit history, potentially affecting their credit score positively.   **SELECT**  **CASE**  **WHEN** **ABS**(**FLOOR**(DAYS\_BIRTH / 365)) **BETWEEN** 20 **AND** 29 **THEN** **'20-29'**  **WHEN** **ABS**(**FLOOR**(DAYS\_BIRTH / 365)) **BETWEEN** 30 **AND** 39 **THEN** **'30-39'**  **WHEN** **ABS**(**FLOOR**(DAYS\_BIRTH / 365)) **BETWEEN** 40 **AND** 49 **THEN** **'40-49'**  **WHEN** **ABS**(**FLOOR**(DAYS\_BIRTH / 365)) **BETWEEN** 50 **AND** 59 **THEN** **'50-59'**  **WHEN** **ABS**(**FLOOR**(DAYS\_BIRTH / 365)) >= 60 **THEN** **'60+'**  **ELSE** **'Under 20'**  **END** **AS** *Age\_Category*,  **COUNT**(DAYS\_BIRTH) **AS** *Applicant\_Count*,  **ROUND**(**AVG**(EXT\_SOURCE\_1), 2) **AS** *Avg\_Ext\_Source\_1*,  **ROUND**(**AVG**(EXT\_SOURCE\_2), 2) **AS** *Avg\_Ext\_Source\_2*,  **ROUND**(**AVG**(EXT\_SOURCE\_3), 2) **AS** *Avg\_Ext\_Source\_3*  **FROM**  application  **WHERE**  DAYS\_BIRTH **IS** **NOT** **NULL** **AND**  EXT\_SOURCE\_1 **IS** **NOT** **NULL** **AND**  EXT\_SOURCE\_2 **IS** **NOT** **NULL** **AND**  EXT\_SOURCE\_3 **IS** **NOT** **NULL**  **GROUP** **BY**  *Age\_Category*  **ORDER** **BY**  *Age\_Category*;  **Output:**   **Observation:**Age Category: 20-29  1. **Total Applicant:** 15,712 2. **Average** **EXT\_SOURCE\_1**: 0.32 3. **Average EXT\_SOURCE\_2**: 0.47 4. **Average EXT\_SOURCE\_3**: 0.42   **Analysis:**   * **Credit Profile**: The low average scores indicate a lack of established credit history. This demographic is often just starting their financial journeys, possibly having limited experience with credit products. * **Behavioral Traits**: Younger individuals may have less stable employment and income, leading to a higher likelihood of payment difficulties. This can stem from lifestyle changes, such as job transitions or further education.  **Implications for Lenders:**  * **Risk Assessment**: Higher perceived risk due to lower credit scores. Lenders may want to implement stricter criteria or require co-signers for loans. * **Educational Opportunities**: Introducing financial education initiatives can help improve credit scores among younger borrowers, ultimately benefiting both the clients and the lenders.  **Recommendations:**  * Conduct workshops or webinars on credit management and financial planning tailored to this age group. * Consider offering small, low-interest loans that can help build credit without imposing overwhelming debt.  Age Category: 30-39  1. **Total Applicant:** 38,487 2. **Average** **EXT\_SOURCE\_1**: 0.43 3. **Average EXT\_SOURCE\_2**: 0.52 4. **Average EXT\_SOURCE\_3**: 0.49   **Analysis:**   * **Maturity and Stability**: This group is likely more settled in their careers and personal lives, contributing to improved credit scores compared to the younger cohort. * **Loan Purpose**: Commonly applying for loans related to significant life events such as buying a home or funding education, indicating both need and opportunity for larger credit amounts.   **Implications for Lenders:**   * **Loan Customization**: Tailoring products for home purchases, family expansions, or educational loans can align with the needs of this demographic. * **Flexible Terms**: Offering flexible repayment plans could appeal to this group, potentially enhancing loan approval rates while reducing defaults.   **Recommendations:**   * Develop targeted marketing strategies that address the specific financial needs of this age group. * Assess the impact of economic fluctuations on this demographic, as they may be more sensitive to changes in job security or interest rates.  Age Category: 40-49  1. **Total Applicant:** 30,036 2. **Average** **EXT\_SOURCE\_1**: 0.55 3. **Average EXT\_SOURCE\_2**: 0.55 4. **Average EXT\_SOURCE\_3**: 0.52   **Analysis:**   * **Improved Credit Scores**: The stability in their financial situations is reflected in the relatively higher scores. Many in this group have built substantial credit histories and may have accumulated assets. * **Changing Needs**: This group might be focusing on home renovations, supporting children’s education, or preparing for retirement.   **Implications for Lenders:**   * **Targeted Loan Products**: Products that cater to home equity loans or education financing could be beneficial. * **Relationship Building**: As many applicants in this group might be looking for long-term financial partnerships, fostering relationships through customer service can enhance loyalty.   **Recommendations:**   * Monitor market trends that could affect this demographic, such as housing market fluctuations or educational costs. * Encourage refinancing options that allow for better interest rates based on their improved creditworthiness.  Age Category: 50-59  1. **Total Applicant:** 18,019 2. **Average** **EXT\_SOURCE\_1**: 0.66 3. **Average EXT\_SOURCE\_2**: 0.55 4. **Average EXT\_SOURCE\_3**: 0.53   **Analysis:**   * **Financial Confidence**: The higher average scores suggest greater financial stability and experience, likely correlating with higher income levels or accumulated wealth. * **Life Stage Considerations**: This group may be looking for loans for retirement planning, investment opportunities, or healthcare expenses.   **Implications for Lenders:**   * **Investment Products**: Offer loans that facilitate investment in retirement accounts or business ventures, as this group may be interested in building wealth. * **Health-Centric Loans**: Products designed for healthcare or long-term care could meet the emerging needs of this demographic.   **Recommendations:**   * Explore partnerships with financial advisors to offer comprehensive financial planning services. * Assess the potential for long-term loans that consider retirement income and lower future repayment burdens.  Age Category: 60+  1. **Total Applicant:** 7,335 2. **Average** **EXT\_SOURCE\_1**: 0.73 3. **Average EXT\_SOURCE\_2**: 0.54 4. **Average EXT\_SOURCE\_3**: 0.56   **Analysis:**   * **Established Credit Histories**: The highest average credit scores indicate that older adults generally have the most stable financial situations, likely benefiting from substantial credit histories. * **Changing Financial Goals**: This demographic may seek loans for downsizing, retirement homes, or consolidating existing debt.   **Implications for Lenders:**   * **Retirement-Focused Products**: Consider offering loans specifically tailored for retirement-related expenses or downsizing. * **Risk Management**: While this group is typically lower risk, lenders should still assess individual financial situations, especially concerning potential healthcare costs.   **Recommendations:**   * Create informational resources or workshops focused on retirement financing and planning. * Develop personalized loan options that respect their fixed income, with emphasis on lower payments or longer terms.   **Conclusion**  The breakdown of loan applicants by age category reveals distinct patterns in creditworthiness, financial needs, and potential risks. By understanding these nuances, lenders can better tailor their products and services, effectively meet the needs of each demographic, and mitigate risks associated with loan defaults. Continuous analysis and adjustments based on changing economic conditions will be crucial in maintaining relevant and effective lending strategies.   * **INCOME (AMT\_INCOME\_TOTAL):**   **SELECT**  *Income\_Category*,  **COUNT**(\*) **AS** *Total\_Applicant*,  **ROUND**(**AVG**(EXT\_SOURCE\_1), 2) **AS** *Avg\_Ext\_Source\_1*,  **ROUND**(**AVG**(EXT\_SOURCE\_2), 2) **AS** *Avg\_Ext\_Source\_2*,  **ROUND**(**AVG**(EXT\_SOURCE\_3), 2) **AS** *Avg\_Ext\_Source\_3*  **FROM** (  **SELECT**  **CASE**  **WHEN** AMT\_INCOME\_TOTAL **IS** **NULL** **THEN** **'Income Not Provided'**  **WHEN** AMT\_INCOME\_TOTAL < 15000 **THEN** **'Very Low Income'**  **WHEN** AMT\_INCOME\_TOTAL **BETWEEN** 15000 **AND** 29999 **THEN** **'Low Income'**  **WHEN** AMT\_INCOME\_TOTAL **BETWEEN** 30000 **AND** 49999 **THEN** **'Lower-Medium Income'**  **WHEN** AMT\_INCOME\_TOTAL **BETWEEN** 50000 **AND** 74999 **THEN** **'Upper-Medium Income'**  **WHEN** AMT\_INCOME\_TOTAL **BETWEEN** 75000 **AND** 99999 **THEN** **'High Income'**  **ELSE** **'Very High Income'**  **END** **AS** *Income\_Category*,  EXT\_SOURCE\_1,  EXT\_SOURCE\_2,  EXT\_SOURCE\_3,  TARGET  **FROM**  application  **WHERE**  EXT\_SOURCE\_1 **IS** **NOT** **NULL** **AND**  EXT\_SOURCE\_2 **IS** **NOT** **NULL** **AND**  EXT\_SOURCE\_3 **IS** **NOT** **NULL**  ) **AS** *IncomeData*  **GROUP** **BY**  *Income\_Category*  **ORDER** **BY**  **CASE**  **WHEN** Income\_Category = **'Income Not Provided'** **THEN** 1  **WHEN** Income\_Category = **'Very Low Income'** **THEN** 2  **WHEN** Income\_Category = **'Low Income'** **THEN** 3  **WHEN** Income\_Category = **'Lower-Medium Income'** **THEN** 4  **WHEN** Income\_Category = **'Upper-Medium Income'** **THEN** 5  **WHEN** Income\_Category = **'High Income'** **THEN** 6  **ELSE** 7  **END**;  **Output:**    **Observation:**   1. **Low Income** 2. **Total Applicants**: 23 3. **Average External Source 1**: 0.55 4. **Average External Source 2**: 0.49 5. **Average External Source 3**: 0.53  * **Insights**: * **Engagement Level**: The very low number of applicants in this category (23) may indicate barriers to loan acquisition for low-income individuals, such as stricter eligibility criteria or a lack of awareness about available loans. * **Creditworthiness**: Despite low income, the average external source scores are relatively high. This suggests that those who do apply may have maintained good financial behavior and credit management practices, potentially due to limited access to credit options. * **Risk Perception**: Lenders might perceive this group as higher risk due to their low income, but the average scores indicate that not all low-income applicants are necessarily credit-risky.  1. **Lower-Medium Income**  * **Total Applicants**: 870 * **Average External Source 1**: 0.55 * **Average External Source 2**: 0.47 * **Average External Source 3**: 0.57 * **Insights**: * **Increased Participation**: The jump in the number of applicants to 870 indicates that individuals in this income bracket are more likely to engage with credit products. This could reflect improved financial literacy or a greater need for loans. * **Credit Scores**: The slightly lower average for External Source 2 (0.47) compared to the other sources suggests potential issues with this group’s credit history or payment behaviors. It raises the question of whether this category is starting to experience financial stress as they balance needs against income constraints. * **Diversity in Financial Profiles**: The variation in external source scores hints at a mix of applicants; some may be managing well while others struggle, reflecting the challenges of navigating credit in this income range.  1. **Upper-Medium Income**  * **Total Applicants**: 4,918 * **Average External Source 1**: 0.52 * **Average External Source 2**: 0.48 * **Average External Source 3**: 0.54 * **Insights**: * **Strong Engagement**: With nearly 5,000 applicants, this category shows substantial engagement with credit products, indicating that individuals are likely pursuing loans for significant expenditures (e.g., home purchases, education). * **Stable Credit Behavior**: The average scores are slightly lower than in the previous category, suggesting that as income increases, applicants may take on more financial risks or complexities, leading to a diverse credit landscape. * **Strategic Borrowing**: This group may be leveraging credit to optimize their financial situations, such as investing in assets, indicating a shift from mere necessity to strategic financial planning.  1. **High Income**  * **Total Applicants**: 12,100 * **Average External Source 1**: 0.50 * **Average External Source 2**: 0.49 * **Average External Source 3**: 0.51 * **Insights**: * **High Volume Participation**: The high number of applicants (12,100) reflects confidence in taking on loans, possibly due to higher disposable income and a proactive approach to financing. * **Credit Management Dynamics**: The average external scores indicate that this group may be balancing larger debts, potentially for investments or lifestyle choices. The consistent scores across the sources suggest that while they may have the means to manage debts, they are possibly facing increased scrutiny from lenders. * **Behavioral Patterns**: High-income applicants might also exhibit different borrowing behaviors, such as preferring larger, more complex financial products, which can influence their credit profiles.  1. **Very High Income**  * **Total Applicants**: 91,678 * **Average External Source 1**: 0.51 * **Average External Source 2**: 0.54 * **Average External Source 3**: 0.49 * **Insights**: * **Dominance in Numbers**: This category has the largest number of applicants, suggesting that financial institutions target this demographic due to their higher purchasing power and greater likelihood of loan repayment. * **Score Variability**: While the external scores are relatively stable, the drop in Source 3 (0.49) indicates potential areas for concern. It might suggest that despite having significant incomes, this group faces specific challenges that can impact credit, such as increased investment activities or multiple debt obligations. * **Implications for Lenders**: The data indicates a need for lenders to tailor their products and risk assessments, as high income does not automatically equate to low risk. Understanding the complexities of this demographic is crucial for effective credit evaluation.   **Overall Analysis and Recommendations:**   * **Trends Across Income Levels**: There is a clear trend showing increased engagement with credit as income rises. However, creditworthiness does not increase linearly, as seen in the fluctuating external scores. * **Risk Management**: Lenders should adopt a nuanced approach to risk assessment, considering both income levels and credit behaviors rather than relying solely on income as a predictor of creditworthiness. * **Targeted Financial Education**: Institutions could enhance financial literacy programs, particularly for lower and lower-medium income groups, to empower more individuals to engage with credit responsibly. * **Product Customization**: Financial products should be tailored to meet the specific needs of each income category, recognizing the diverse motivations and challenges faced by borrowers. |