</talentlabs>

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

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
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	307511
bureau	1716428

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 vital part of someone's financial life, while it dictates the ability to acquire loans, get good interest rates, and even whether one can afford housing or employment. Thus, maintaining a good credit score is important for long-term financial health.

Task 3 Understand Credit Amount and Annuity

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

Credit Amount	Credit amount is the total amount that the borrower can borrow, ranging in different financial instruments like loans, credit cards, and lines of credit. It is determined by the lender based on the borrower's creditworthiness, income, and the type of financial product involved.

Annuity	An annuity is a financial instrument that provides a somewhat predictable income stream, which can be applied to retirement or other long-term financial needs. While an annuity loan is the payment method in which one repays a loan with periodic payments over a certain time, using fixed payments. These repayments are typically comprised of interest and principal, although the interest portion will gradually decline as the balance amount goes down. This is quite similar to how mortgages, personal loans, and auto loans work.

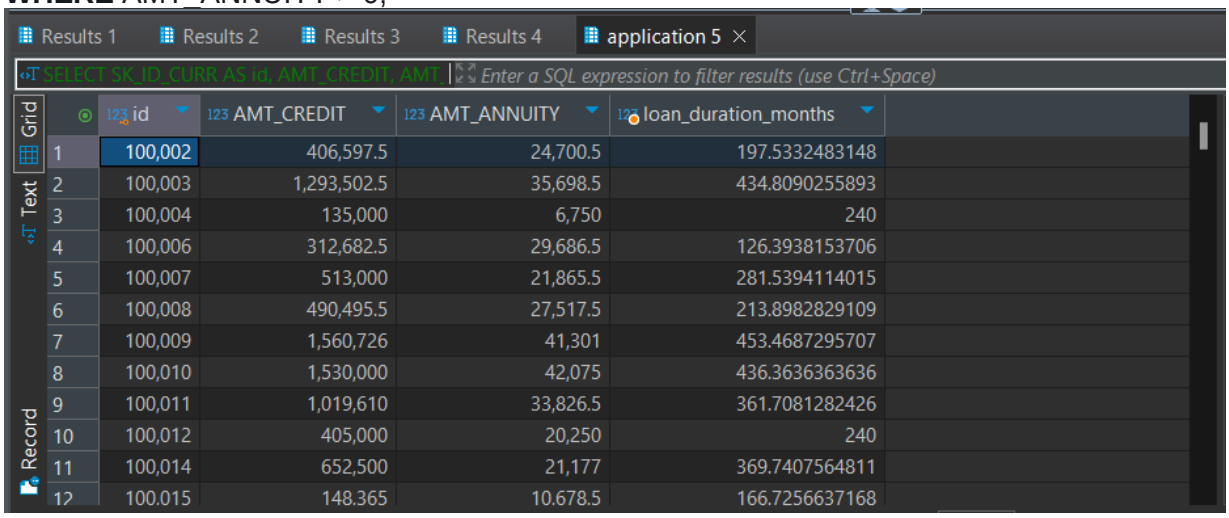
Task 4 Deduce the Loan Duration

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:

We can deduce the loan duration by using the credit amount divided by the loan annuity. Then we further divide it into 12 to get the months durations.

```
SELECT
    SK_ID_CURR AS id,
    AMT_CREDIT,
    AMT_ANNUITY,
    AMT_CREDIT / (AMT_ANNUITY / 12) AS loan_duration_months
FROM application
WHERE AMT_ANNUITY > 0;
```



	id	AMT_CREDIT	AMT_ANNUITY	loan_duration_months
1	100,002	406,597.5	24,700.5	197.5332483148
2	100,003	1,293,502.5	35,698.5	434.8090255893
3	100,004	135,000	6,750	240
4	100,006	312,682.5	29,686.5	126.3938153706
5	100,007	513,000	21,865.5	281.5394114015
6	100,008	490,495.5	27,517.5	213.8982829109
7	100,009	1,560,726	41,301	453.4687295707
8	100,010	1,530,000	42,075	436.3636363636
9	100,011	1,019,610	33,826.5	361.7081282426
10	100,012	405,000	20,250	240
11	100,014	652,500	21,177	369.7407564811
12	100,015	148,365	10,678.5	166.7256637168

</talentlabs>

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

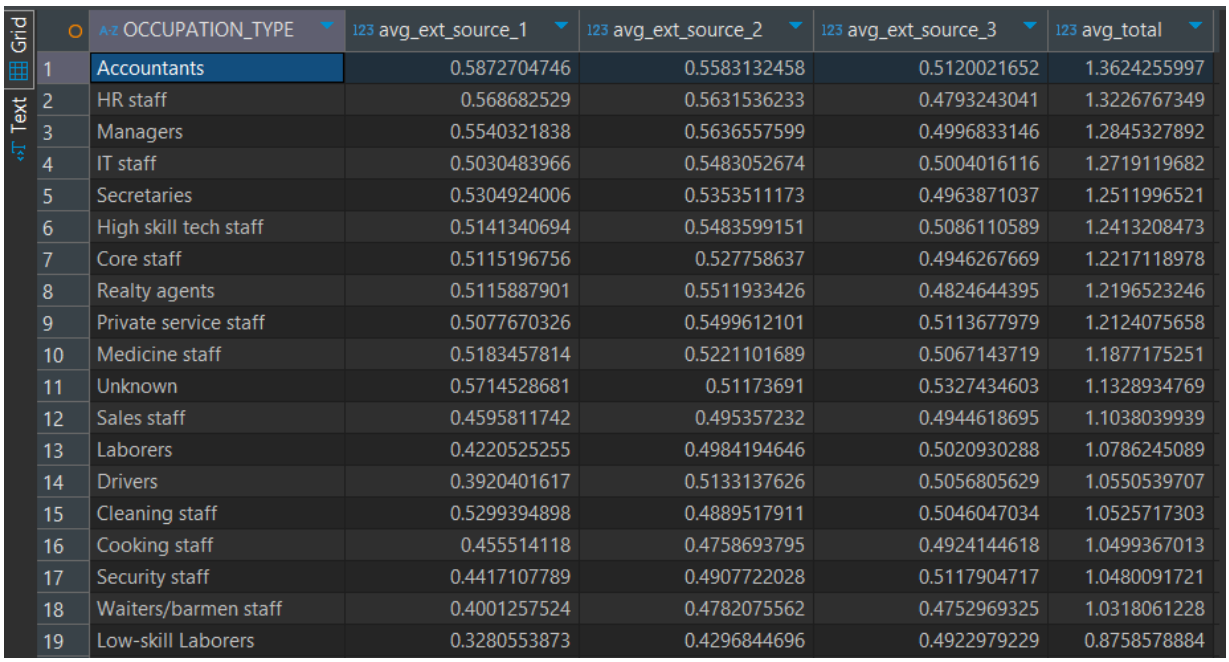
In the “application” table try to explore if there are any columns affecting the credit score. For example, is gender a factor?

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. Occupation type affecting Credit Score.

```
SELECT
   CASE
      WHEN OCCUPATION_TYPE IS NULL OR OCCUPATION_TYPE = '' THEN
'Unknown'
      ELSE OCCUPATION_TYPE
   END AS OCCUPATION_TYPE,
   AVG(EXT_SOURCE_1) AS avg_ext_source_1,
   AVG(EXT_SOURCE_2) AS avg_ext_source_2,
   AVG(EXT_SOURCE_3) AS avg_ext_source_3,
   AVG(COALESCE(EXT_SOURCE_1, 0) + COALESCE(EXT_SOURCE_2, 0) +
COALESCE(EXT_SOURCE_3, 0)) AS avg_total
FROM
   application
GROUP BY
   CASE
      WHEN OCCUPATION_TYPE IS NULL OR OCCUPATION_TYPE = '' THEN
'Unknown'
      ELSE OCCUPATION_TYPE
   END
ORDER BY
   avg_total DESC;
```

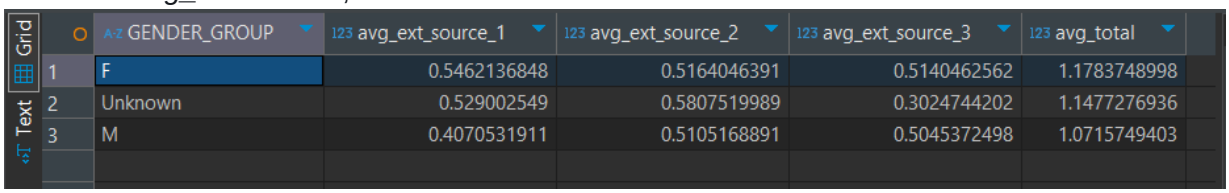


	OCCUPATION_TYPE	avg_ext_source_1	avg_ext_source_2	avg_ext_source_3	avg_total
1	Accountants	0.5872704746	0.5583132458	0.5120021652	1.3624255997
2	HR staff	0.568682529	0.5631536233	0.4793243041	1.3226767349
3	Managers	0.5540321838	0.5636557599	0.4996833146	1.2845327892
4	IT staff	0.5030483966	0.5483052674	0.5004016116	1.2719119682
5	Secretaries	0.5304924006	0.5353511173	0.4963871037	1.2511996521
6	High skill tech staff	0.5141340694	0.5483599151	0.5086110589	1.2413208473
7	Core staff	0.5115196756	0.527758637	0.4946267669	1.2217118978
8	Realty agents	0.5115887901	0.5511933426	0.4824644395	1.2196523246
9	Private service staff	0.5077670326	0.5499612101	0.5113677979	1.2124075658
10	Medicine staff	0.5183457814	0.5221101689	0.5067143719	1.1877175251
11	Unknown	0.5714528681	0.51173691	0.5327434603	1.1328934769
12	Sales staff	0.4595811742	0.495357232	0.4944618695	1.1038039939
13	Laborers	0.4220525255	0.4984194646	0.5020930288	1.0786245089
14	Drivers	0.3920401617	0.5133137626	0.5056805629	1.0550539707
15	Cleaning staff	0.5299394898	0.4889517911	0.5046047034	1.0525717303
16	Cooking staff	0.455514118	0.4758693795	0.4924144618	1.0499367013
17	Security staff	0.4417107789	0.4907722028	0.5117904717	1.0480091721
18	Waiters/barmen staff	0.4001257524	0.4782075562	0.4752969325	1.0318061228
19	Low-skill Laborers	0.3280553873	0.4296844696	0.4922979229	0.8758578884

We can see that certain occupations have a much better credit score compared to others.

2. Gender affecting Credit Score

```
SELECT
    CASE
        WHEN CODE_GENDER = 'XNA' THEN 'Unknown'
        ELSE CODE_GENDER
    END AS GENDER_GROUP,
    AVG(EXT_SOURCE_1) AS avg_ext_source_1,
    AVG(EXT_SOURCE_2) AS avg_ext_source_2,
    AVG(EXT_SOURCE_3) AS avg_ext_source_3,
    AVG(COALESCE(EXT_SOURCE_1, 0) + COALESCE(EXT_SOURCE_2, 0) +
COALESCE(EXT_SOURCE_3, 0)) AS avg_total
FROM
    application
GROUP BY
    GENDER_GROUP
ORDER BY
    avg_total DESC;
```



	GENDER_GROUP	avg_ext_source_1	avg_ext_source_2	avg_ext_source_3	avg_total
1	F	0.5462136848	0.5164046391	0.5140462562	1.1783748998
2	Unknown	0.529002549	0.5807519989	0.3024744202	1.1477276936
3	M	0.4070531911	0.5105168891	0.5045372498	1.0715749403

We can see that female have a better credit score compared to male.

3. Education Level affecting Credit Score

```
SELECT
   CASE
      WHEN NAME_EDUCATION_TYPE is null or NAME_EDUCATION_TYPE = ' '
THEN 'Unknown'
      ELSE NAME_EDUCATION_TYPE
   END AS Education_Level,
   AVG(EXT_SOURCE_1) AS avg_ext_source_1,
   AVG(EXT_SOURCE_2) AS avg_ext_source_2,
   AVG(EXT_SOURCE_3) AS avg_ext_source_3,
   AVG(COALESCE(EXT_SOURCE_1, 0) + COALESCE(EXT_SOURCE_2, 0) +
COALESCE(EXT_SOURCE_3, 0)) AS avg_total
FROM
   application
GROUP BY
   Education_level
ORDER BY
   avg_total DESC;
```

	Education_Level	avg_ext_source_1	avg_ext_source_2	avg_ext_source_3	avg_total
1	Higher education	0.5451198929	0.5563456421	0.5033297503	1.272163444
2	Academic degree	0.5540810216	0.5668482812	0.4849975007	1.265931857
3	Incomplete higher	0.4330224521	0.5099306096	0.4587957097	1.109795473
4	Secondary / secondary special	0.48708644	0.5013348702	0.5156793974	1.101817561
5	Lower secondary	0.4400374791	0.4480449672	0.5266361704	0.960385862

We can see that having a higher education can have a better credit score compared to lower education level.

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

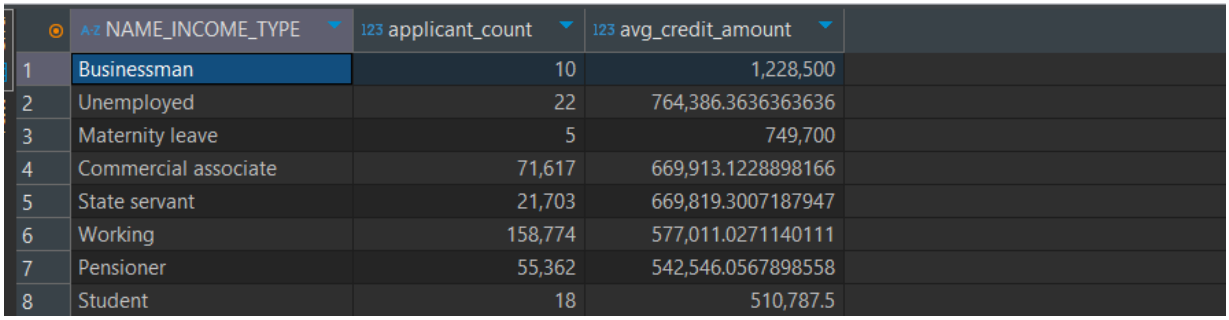
Who is going to lend more money than others? In this task, we want to see are there any factors affecting the credit amount. **Do the analysis of at least 3 factors**

Please explain your findings with SQL statements and results:

1. Income type affecting credit amount

```
SELECT
   NAME_INCOME_TYPE,
   COUNT(*) AS applicant_count,
   AVG(AMT_CREDIT) AS avg_credit_amount
FROM
   application
```

```
GROUP BY
    NAME_INCOME_TYPE
ORDER BY
    avg_credit_amount DESC;
```

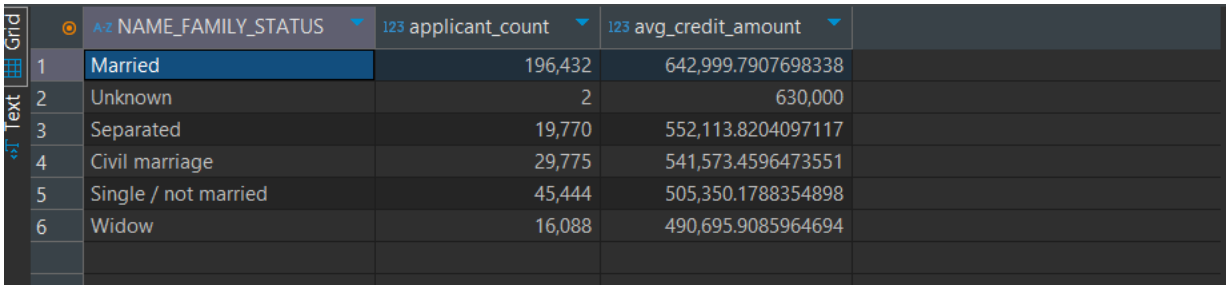


	NAME_INCOME_TYPE	applicant_count	avg_credit_amount
1	Businessman	10	1,228,500
2	Unemployed	22	764,386.3636363636
3	Maternity leave	5	749,700
4	Commercial associate	71,617	669,913.1228898166
5	State servant	21,703	669,819.3007187947
6	Working	158,774	577,011.0271140111
7	Pensioner	55,362	542,546.0567898558
8	Student	18	510,787.5

We can see the type to loan more is businessman and unemployed where businessman loan more probably for their business or investment while unemployed loan more for their survival.

2. Family status affecting credit amount

```
SELECT
    NAME_FAMILY_STATUS ,
    COUNT(*) AS applicant_count,
    AVG(AMT_CREDIT) AS avg_credit_amount
FROM
    application
GROUP BY
    NAME_FAMILY_STATUS
ORDER BY
    avg_credit_amount DESC;
```



	NAME_FAMILY_STATUS	applicant_count	avg_credit_amount
1	Married	196,432	642,999.7907698338
2	Unknown	2	630,000
3	Separated	19,770	552,113.8204097117
4	Civil marriage	29,775	541,573.4596473551
5	Single / not married	45,444	505,350.1788354898
6	Widow	16,088	490,695.9085964694

Married people have a higher credit amount probably for housing and other necessities.

3. Number of children affecting credit amount

```
SELECT
    CNT_CHILDREN ,
    COUNT(*) AS applicant_count,
    AVG(AMT_CREDIT) AS avg_credit_amount
FROM
    application
GROUP BY
```

```
    CNT_CHILDREN
ORDER BY
    avg_credit_amount DESC;
```

	CNT_CHILDREN	applicant_count	avg_credit_amount
1	11	1	1,007,761.5
2	14	3	833,070
3	4	429	618,222.6713286713
4	3	3,717	607,895.6162227603
5	2	26,749	604,827.0718344611
6	0	215,371	599,103.0428330648
7	1	61,119	595,640.2978942718
8	9	2	583,285.5
9	5	84	571,460.8928571428
10	6	21	523,736.1428571428
11	8	2	497,520
12	12	2	474,390
13	7	7	457,944.4285714286
14	19	2	202,275

We can see there a correlation for number of children and amount of credit where beside the outlier (19, 14, 12, 11) of number children, it show a trend of increasing from 0 to 4 children.

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:

First we have to determine how we want to calculate payment difficulties. We can calculate the payment difficulties by adding the total target which is 1 and 0 and divide it by the count of applicant. Thus, we can get the average of it.

1. Income type affecting payment difficulties

```
SELECT
    NAME_INCOME_TYPE,
    COUNT(*) AS total_applicants,
    SUM(CASE WHEN TARGET = 1 THEN 1 ELSE 0 END) AS payment_difficulties,
    AVG(TARGET) * 100 AS difficulty_rate
FROM
    application
GROUP BY
```

```
    NAME_INCOME_TYPE
ORDER BY
    difficulty_rate DESC;
```

	NAME_INCOME_TYPE	total_applicants	payment_difficulties	difficulty_rate
1	Maternity leave	5	2	40
2	Unemployed	22	8	36.3636
3	Working	158,774	15,224	9.5885
4	Commercial associate	71,617	5,360	7.4843
5	State servant	21,703	1,249	5.755
6	Pensioner	55,362	2,982	5.3864
7	Student	18	0	0
8	Businessman	10	0	0

We can see that on maternity leave and unemployed have a higher difficulty rate compared others.

2. Education level affecting payment difficulties

```
SELECT
    NAME_EDUCATION_TYPE,
    COUNT(*) AS total_applicants,
    SUM(CASE WHEN TARGET = 1 THEN 1 ELSE 0 END) AS payment_difficulties,
    AVG(TARGET) * 100 AS difficulty_rate
FROM
    application
GROUP BY
    NAME_EDUCATION_TYPE
ORDER BY
    difficulty_rate DESC;
```

	NAME_EDUCATION_TYPE	total_applicants	payment_difficulties	difficulty_rate
1	Lower secondary	3,816	417	10.9277
2	Secondary / secondary special	218,391	19,524	8.9399
3	Incomplete higher	10,277	872	8.485
4	Higher education	74,863	4,009	5.3551
5	Academic degree	164	3	1.8293

We can see that a higher education level can influence the payment difficulties with a lower education level can make it hard for payment. This can be caused by income based on education and discipline in payment of loans.

3. Occupation type affecting payment difficulties

```
SELECT
    COALESCE(NULLIF(OCCUPATION_TYPE, ''), 'Unknown') AS
OCCUPATION_TYPE,
    COUNT(*) AS total_applicants,
```

```
    SUM(CASE WHEN TARGET = 1 THEN 1 ELSE 0 END) AS payment_difficulties,
    AVG(TARGET) * 100 AS difficulty_rate
FROM
    application
GROUP BY
    COALESCE(NULLIF(OCCUPATION_TYPE, ''), 'Unknown')
ORDER BY
    difficulty_rate DESC;
```

	OCCUPATION_TYPE	total_applicants	payment_difficulties	difficulty_rate
1	Low-skill Laborers	2,093	359	17.1524
2	Drivers	18,603	2,107	11.3261
3	Waiters/barmen staff	1,348	152	11.276
4	Security staff	6,721	722	10.7424
5	Laborers	55,186	5,838	10.5788
6	Cooking staff	5,946	621	10.444
7	Sales staff	32,102	3,092	9.6318
8	Cleaning staff	4,653	447	9.6067
9	Realty agents	751	59	7.8562
10	Secretaries	1,305	92	7.0498
11	Medicine staff	8,537	572	6.7002
12	Private service staff	2,652	175	6.5988
13	Unknown	96,391	6,278	6.5131
14	IT staff	526	34	6.4639

We can see that low skill laborers and driver have a higher difficulty for payment compare to others.

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

</talentlabs>

Task 7 Is the number of other loans affecting the payment difficulties?

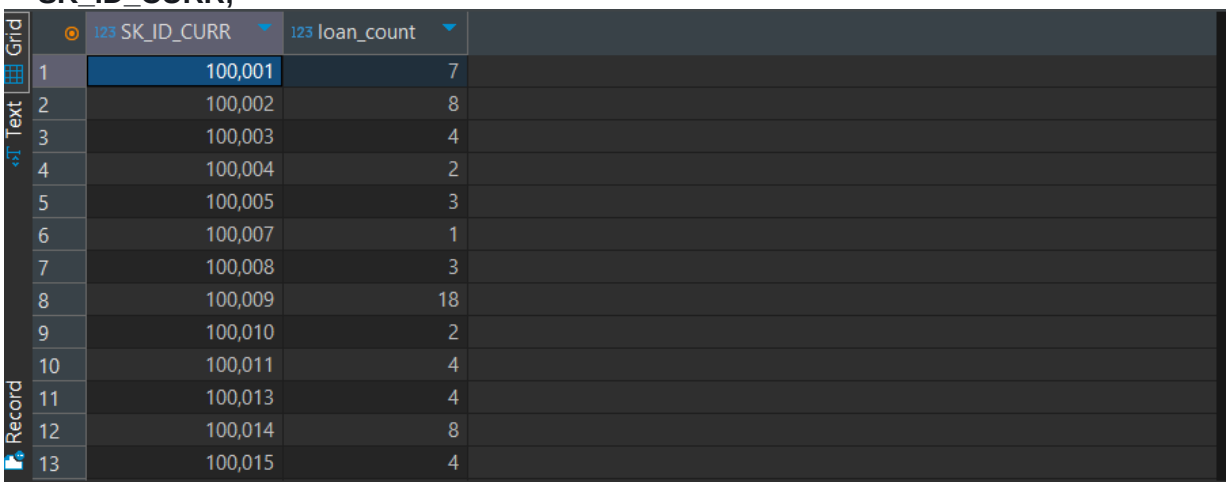
We want to see if loan applicants have other historical loans affecting their payment abilities.
Hints:

- 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:

1. First, we have to count the loan amount for each applicant

```
SELECT
   SK_ID_CURR,
   COUNT(*) AS loan_count
FROM
   bureau
GROUP BY
   SK_ID_CURR;
```

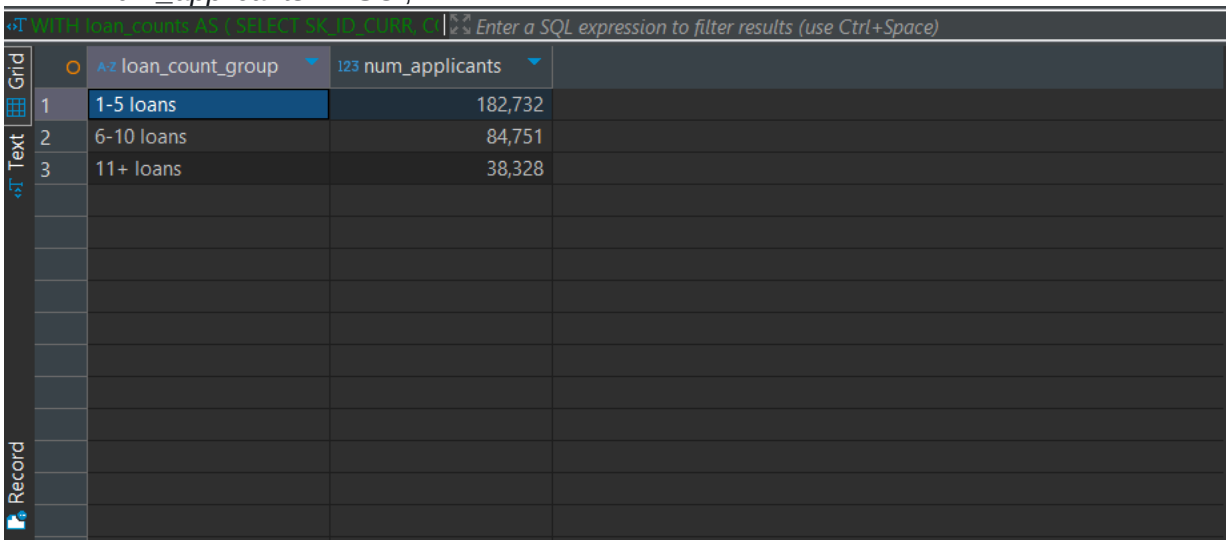


	SK_ID_CURR	loan_count
1	100,001	7
2	100,002	8
3	100,003	4
4	100,004	2
5	100,005	3
6	100,007	1
7	100,008	3
8	100,009	18
9	100,010	2
10	100,011	4
11	100,013	4
12	100,014	8
13	100,015	4

2. Then we group the loan count into certain groups (0, 1-5, 6-10, 11+)

```
WITH loan_counts AS (
   SELECT
      SK_ID_CURR,
      COUNT(*) AS loan_count
   FROM
      bureau
   GROUP BY
      SK_ID_CURR
)
SELECT
   loan_count_group,
   COUNT(*) AS num_applicants
FROM (
   SELECT
      SK_ID_CURR,
      CASE
```

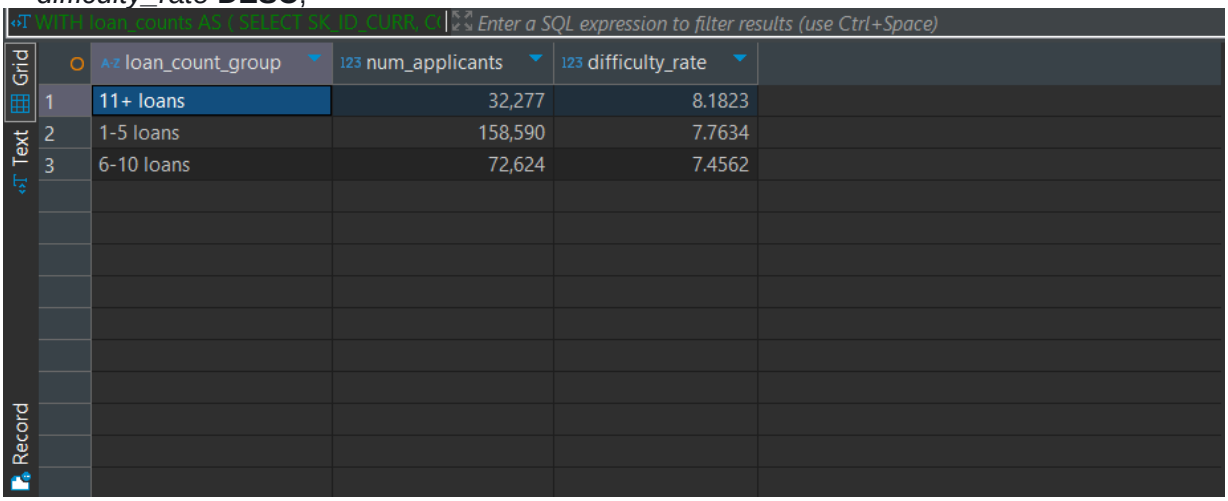
```
        WHEN loan_count = 0 THEN '0 loans'
        WHEN loan_count BETWEEN 1 AND 5 THEN '1-5 loans'
        WHEN loan_count BETWEEN 6 AND 10 THEN '6-10 loans'
        ELSE '11+ loans'
    END AS loan_count_group
  FROM
    loan_counts
) grouped_loans
GROUP BY
  loan_count_group
ORDER BY
    num_applicants DESC ;
```



3. Then we can see the relation between average loan count and the target to see their relation.

```
WITH loan_counts AS (
  SELECT
    SK_ID_CURR,
    COUNT(*) AS loan_count
  FROM
    bureau
  GROUP BY
    SK_ID_CURR
),
loan_groups AS (
  SELECT
    SK_ID_CURR,
    CASE
      WHEN loan_count = 0 THEN '0 loans'
      WHEN loan_count BETWEEN 1 AND 5 THEN '1-5 loans'
      WHEN loan_count BETWEEN 6 AND 10 THEN '6-10 loans'
      ELSE '11+ loans'
    END AS loan_count_group
  FROM
    loan_counts
```

```
)
SELECT
    g.loan_count_group,
    COUNT(*) AS num_applicants,
    AVG(app.TARGET) * 100 AS difficulty_rate
FROM
    loan_groups g
JOIN
    application app
ON
    g.SK_ID_CURR = app.SK_ID_CURR
GROUP BY
    g.loan_count_group
ORDER BY
    difficulty_rate DESC;
```



	loan_count_group	num_applicants	difficulty_rate
1	11+ loans	32,277	8.1823
2	1-5 loans	158,590	7.7634
3	6-10 loans	72,624	7.4562

Based on our data we can see that applicants with more loans have a higher difficulty rate. However, for applicant between 1 – 5 loans and 6 – 10 loans the difficulty rate is comparable.

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

a) Relationship between education type and days employed with credit score and payment difficulty.

```
SELECT
    TARGET,
    DAYS_EMPLOYED,
    NAME_EDUCATION_TYPE,
    AMT_CREDIT,
```

```
      (COALESCE(a.EXT_SOURCE_1, 0) + COALESCE(a.EXT_SOURCE_2, 0) +
COALESCE(a.EXT_SOURCE_3, 0)) /
   NULLIF(
      (CASE WHEN a.EXT_SOURCE_1 IS NOT NULL THEN 1 ELSE 0 END +
       CASE WHEN a.EXT_SOURCE_2 IS NOT NULL THEN 1 ELSE 0 END +
       CASE WHEN a.EXT_SOURCE_3 IS NOT NULL THEN 1 ELSE 0 END),
      0
   ) AS avg_scores FROM
   application AS Applicant
WHERE
   TARGET = 1
ORDER BY
   AMT_CREDIT DESC;
```

	TARGET	DAYS_EMPLOYED	NAME_EDUCATION_TYPE	AMT_CREDIT	avg_scores
1	1	-1,455	Secondary / secondary special	4,027,680	0.1646373542
2	1	-1,379	Secondary / secondary special	3,020,760	0.4978208368
3	1	-1,445	Secondary / secondary special	2,961,000	0.1820744437
4	1	-1,229	Secondary / secondary special	2,695,500	0.2902265044
5	1	-1,474	Higher education	2,695,500	0.5603692797
6	1	-530	Secondary / secondary special	2,695,500	0.4645460921
7	1	-205	Incomplete higher	2,695,500	0.2758561548
8	1	-2,186	Higher education	2,687,355	0.2230675545
9	1	-1,084	Higher education	2,606,400	0.5072601541
10	1	-169	Higher education	2,575,125	0.5723746542
11	1	-2,629	Incomplete higher	2,517,300	0.5572099321
12	1	-3,379	Higher education	2,517,300	0.2642526499
13	1	-1,354	Higher education	2,517,300	0.2212024422
14	1	-4,795	Higher education	2,517,300	0.3252935096

We can see here that two factors affected the credit scores and the payment difficulty. Firstly, the education type where the top applicant highest credit amount are people that only secondary school, incomplete higher education and higher education with the top 14 don't have any degree level education. Next, days employed play a major role in detecting the credit scores and payment difficulty. When applying for the loan, the top applicants all have negative days employed meaning that they were not employed when applying for the loan. That is one of the main reasons for payment difficulty.

b) Relationship between type of credit and sum of debt with credit scores and payment difficulty.

```
SELECT
   b.SK_ID_CURR,
   b.CREDIT_TYPE,
   SUM(b.AMT_CREDIT_SUM_DEBT) AS Total_Debt,
   a.TARGET,
   (COALESCE(a.EXT_SOURCE_1, 0) + COALESCE(a.EXT_SOURCE_2, 0) +
COALESCE(a.EXT_SOURCE_3, 0)) /
   NULLIF(
      (CASE WHEN a.EXT_SOURCE_1 IS NOT NULL THEN 1 ELSE 0 END +
       CASE WHEN a.EXT_SOURCE_2 IS NOT NULL THEN 1 ELSE 0 END +
       CASE WHEN a.EXT_SOURCE_3 IS NOT NULL THEN 1 ELSE 0 END),
```

```
        0
    ) AS avg_scores
FROM
    bureau AS b
JOIN
    application AS a
ON
    b.SK_ID_CURR = a.SK_ID_CURR
WHERE
    a.TARGET = 1
GROUP BY
    b.SK_ID_CURR,
    b.CREDIT_TYPE,
    a.TARGET
ORDER BY
    Total_Debt DESC;
```

	SK_ID_CURR	CREDIT_TYPE	Total_Debt	TARGET	avg_scores
1	263,928	Mortgage	35,108,518.5	1	0.1693162822
2	443,521	Mortgage	32,533,254	1	0.2644170853
3	421,650	Consumer credit	25,620,975	1	0.3963789748
4	261,295	Consumer credit	22,321,781.325	1	0.5708155193
5	158,231	Mortgage	21,947,427	1	0.3000234885
6	249,882	Mortgage	21,417,165	1	0.6095228059
7	300,710	Mortgage	18,764,883	1	0.3803394658
8	363,442	Mortgage	17,748,769.5	1	0.2412101132
9	418,171	Consumer credit	16,245,486	1	0.4797874281
10	449,691	Mortgage	16,053,264	1	0.2348072868
11	114,843	Mortgage	15,972,336	1	0.4333983621
12	127,302	Mortgage	14,966,743.5	1	0.4718325456
13	189,803	Mortgage	14,293,165.5	1	0.276971494
14	123,857	Mortgage	14,217,291	1	0.3085992223

We can see that total debt and type of credit affected the credit scores for applicants. With a high total debt and a type of credit mortgage the credit scores of applicants is lower.