Final project (SQL module)

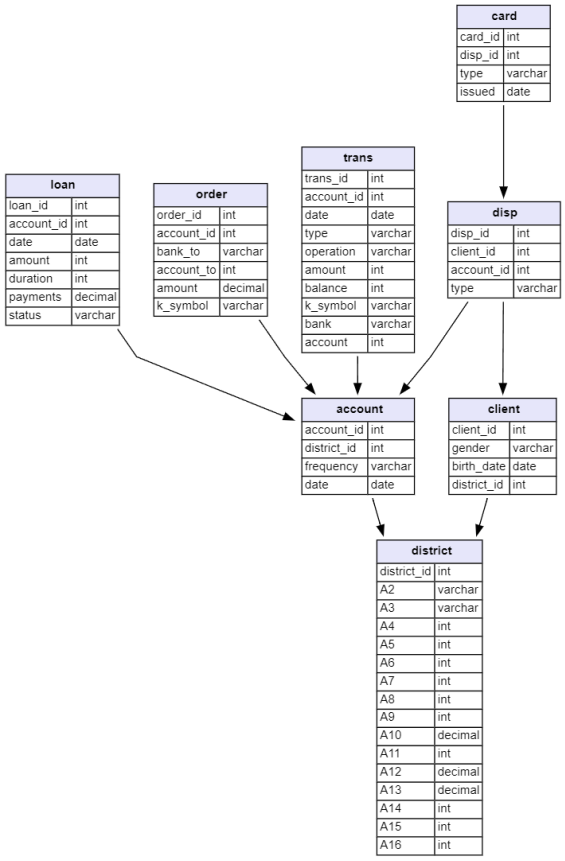
By Veronika Zemcikova

In this project I analyzed a data set which contains a collection of financial information from a Czech bank.

The dataset contains data for more than 5,300 clients with about one million transactions. What's more, the bank has also released data for nearly 700 loans granted and about 900 credit cards issued.

Database structure

- primary keys in individual tables:
  - account – account_id
  - client – client_id
  - card – card_id
  - disp – disp_id
  - district – district_id
  - loan – loan_id
  - order – order_id
  - trans – trans_id
- relationships between particular pairs of tables:
  - account – disp, trans, order, loan – 1 to n
  - disp – card – 1 to n
  - client – disp – 1 to n
  - district – client – 1 to n
  - district – account – 1 to n



In following exercises, the task was to analyze the dataset according to the assignment.

1. History of granted loans
   - using EXTRACT function, I separated parts of the date from the whole date, then used specific aggregate functions and thanks to that I got information from the table necessary to prepare a summary of the granted loans in the required dimensions
   - using a rollup function I displayed the required information as the result of the summary
2. Loan status

- Thanks to the information in the assignment we know that there is a total of 682 granted loans in the database, of which 606 have been repaid and 76 have not
- using the aggregate function COUNT I was able to group loans by their status:
    - A - 203
    - B - 31
    - D - 45
    - C - 403
        - **Loans with the status A and C have been paid (A + C = 606)**

3. Analysis of accounts
    - using aggregate functions I was able to write a query that ranks accounts according to the required criteria
    - as each client has only one loan the loan amount and the average loan amount are the same
    - only fully paid loans were considered
4. Fully paid loans
    - using aggregate function and inner join function to join necessary tables I found out the balance of repaid loans, divided by client gender
    - as a method to check my query I created a CTE table and then compared results from the original query with the result from CTE table
5. Client analysis - part 1
    - modifying the queries from the exercise on repaid loans, I was able to answer the following questions
    - Who has more repaid loans - women or men?
        - Women have more repaid loans (307, men 299).
    - What is the average age of the borrower divided by gender?
        - The average age of female borrowers is 65, the average age of male borrowers is 67.
6. Client analysis - part 2
    - In part 2 I made analysis using temporary table to answer the following questions:
        - Which area has the most clients? – Hl. M. Praha
        - In which area the highest number of loans was paid? – Hl. M. Praha
        - In which area the highest amount of loans was paid? – Hl. M. Praha
7. Client analysis - part 3
    - Using the query created in the previous task and CTE table I determined the percentage of each district in the total amount of loans granted.
8. Selection - part 1
    - Writing a specific query, I checked the database for the clients who met the required results, which were:
        - their account balance is above 1000,
        - they have more than 5 loans,
        - they were born after 1990
    - There were no such clients.
9. Selection - part 2
    - In this exercise I had to make an analysis to determine which condition caused the empty results.

- Writing two different queries I found out that there are no clients who have more than 5 loans and were born after 1990.

10. Expiring cards
    - In this exercise I had to write a query to get data with the following columns:
        - client_id,
        - card_id,
        - expiration_date - assume that the card can be active for 3 years after issue date,
        - client_address (column A3 is enough)
    - Using data functions and inner join I got the necessary information.