



## DATA SET

The tables contain data from a bank. The bank is managing accounts, handling transactions and giving out loans for their clients.

---

The data consists of the following four tables:

- account
- client
- disp
- loan

Below you will find:

- A brief explanation of each table and their columns
  - An ERD (entity relationship diagram) showing the relationship between the tables
- 

## OVERVIEW OF COLUMNS

### ACCOUNT

4500 records x 4 fields

1. account\_id = unique numerical identifier for each bank account in the data set
2. district\_id = numerical identifier for 77 different districts
3. frequency = frequency of issuance of bank statements
4. date = date when the account was created

## **CLIENT**

5369 x 4

1. client\_id = unique numerical identifier for each client in the data set
2. gender = ...
3. birth\_date = ...
4. district\_id = ...

## **DISP**

5369 x 4

The disp (short for disposition) table contains information linking clients with accounts.

1. disp\_id = unique numerical identifier for each relationship between a client and an account in the data set
2. client\_id = ...
3. account\_id = ...
4. type = shows type of relationship between client and account. Either "owner" or "user".

## **LOAN**

682 x 7

1. loan\_id = unique numerical identifier for each loan in the data set
  2. account\_id = ...
  3. date = date when loan was granted
  4. amount = loan amount
  5. duration = loan duration (months)
  6. installment = size of monthly repayment of loan
  7. status = loan status
    - a. A = Finished contract, loan fully repaid on time
    - b. B = Finished contract, loan not fully repaid
    - c. C = Running contract, loan repaid on time (so far)
    - d. D = Running contract, client in debt
-

# ERD

## Entity Relationship Diagram

This is a simplified version of an ERD. The most important information pieces are the **primary keys** and **foreign keys**. A **primary key** is the column (or columns) that contain values that uniquely identify each row in a table. A **foreign key** is what you are using when you are joining **tables**. For instance:

In the **client** table the **primary key** is **client\_id**. In the **disp** table the **primary key** is **disp\_id**. When joining the **disp** table with the **client** table, you will use the **client\_id** column from the **disp** table as the **foreign key**.

```
SELECT * FROM client JOIN disp ON client.client_id = disp.client_id
```

