

Senior Analyst Skills Assessment

Section 1: Loan product and customers

1. Long-Term High-Value Debtors

A) Write a SQL query to identify customers who have been behind schedule for more than 10 days and have an outstanding balance greater than 2000 EUR.

In addition to borrower ID and total outstanding amount include:

- A flag whether their last payment is more than 5% lower than their previous payment
- A flag whether their outstanding balance has increased over the last 3 months.

B) Write a SQL query to find the top 10 customers with the highest total outstanding debt across all loans. Include:

- A breakdown of their total outstanding debt accumulation over the past 6 weeks.
- A column with the total change and % change in their outstanding balance week-over-week.

2. Recovery Rates

A) Write a SQL query to calculate the total daily recovery rate (total payments / total outstanding balance). Also include:

- A 3-week moving average of the recovery rate.

B) Using SQL, calculate the daily recovery rate per borrower and loan to give us better insight into payment behavior over time. Also include:

- A running total of payments and days past due per customer, daily level.
- A flag for days when the recovery rate got worse.

Section 2: Data Analysis and processing

Tasks:

A) Create a forecasting model for daily total repayments. Which technique(s) are you using and how did you choose them? Assume the forecasting range should be 7 days into the future.

B) Split the dataset in half for future A/B testing – each half should have a similar amount of borrowers, loans, loans per country, outstanding debt, ... i.e. make them as equal as possible.

Section 3: Business Insight and Interpretation

Based on your analysis, answer the following:

- Are high outstanding balances more likely to be in delay?
- Are short loan durations more likely to be on time?
- Which country should we focus on?

Instructions & Setup

Instructions:

- Explain any assumptions made while answering the questions.
- Unless explicitly asking for SQL, feel free to use your tooling of choice.
- If time permits, feel free to suggest additional insights based on the data.
- Visualize key insights where applicable

Technical Setup

- Feel free to use any SQL / coding environment you are comfortable with
- In case you have no experience setting up such an environment (not in scope for this assignment), we recommend installing Anaconda and launching a Jupyter Notebook. It comes with native python and you can get your SQL setup going like this:

```
!pip install pandas pandasql
```

```
import pandas as pd
import pandasql as psql

# Load the CSV file into a DataFrame
df = pd.read_csv('your_file.csv')
```

```
# Define your SQL query
query = "SELECT * FROM df WHERE column_name = 'some_value'"

# Execute the query
result = psql.sqldf(query, locals())

# Display the result
print(result)
```

Data dictionary

General info: You received a truncated set of data ordered by report_date_local, containing simulated records for loans **issued** in April, May, and June 2025. Not all loan records are complete. Document issues with data quality as you find them.

borrower_id: Unique identifier of the customer.

loan_id: Unique identifier of the loan.

loan_issued_at: date when loan was issued, i.e. when did the customer get the loan.

country: Country of the loan.

outstanding_balance: Total outstanding balance to be repaid.

repaid_amount_day: Actual (re)payment made to the loan on that date.

days_past_due: Number of days in debt.

report_date_local: Shows the date for which the data is reported.