

Credit Card Default Risk Analysis using SQL

Project Overview

This project focuses on credit card default risk analysis using SQL. The goal is to clean the data, handle inconsistencies, and identify patterns and risk factors associated with customer default. The analysis supports data-driven decision-making for credit risk management.

Datasets Description

- The datasets contains information about 30,000 credit card clients, including:
- Demographics: SEX, EDUCATION, MARRIAGE, AGE
- Payment history: PAY_0 to PAY_6
- Billing and payments amounts: BILL_AMT1 to BILL_AMT6, PAY_AMT1 to PAY_AMT6
- Target Variable: default payment next month(1=default, 0=no default)

Data Cleaning

- No missing values were found in the datasets.
- Inconsistent categorical values were detected:
 - EDUCATION: value like 0,5,6 were mapped to “other” (4)
 - MARRIAGE: value 0 was mapped to “other” (3)
- Created cleaned columns using CASE WHEN logic for further analysis.

```
ALTER TABLE credit_card_data ADD COLUMN EDUCATION_CLEANED INTEGER;  
UPDATE credit_card_data  
SET EDUCATION_CLEANED = CASE  
    WHEN EDUCATION IN (1, 2, 3, 4) THEN EDUCATION  
    ELSE 4  
END;
```

Analysis & key Insights

Overall Default Rate

- Default Rate: 22%

```
SELECT ROUND(100.0 * SUM("default payment next month") / COUNT(*), 2) AS  
default_rate_percent  
FROM credit_card_data;
```

Default Rate by Gender Group

- Male has higher default rate (24%) than female (20%)

Default Rate by Education Group

- Lower education levels correlate with higher default rates.

Default Rate by Marriage Status

- Minor differences observed; “Other” category needs further profiling.

Default Rate by Age Group

- Grouped age into:
 - Under 30
 - 30-39
 - 40-49
 - 50+

Finding:

Under 30 has the highest default rate.

Default Rate by Payment History(PAY_0)

- Clients with delayed payments(PAY_0 >=1) show much higher default risk.

Default Rate by Credit Limit

- The lower credit limit has a higher default rate.

Conclusion

- This project demonstrates how SQL can be used to conduct a complete analysis pipeline.
- Data quality checking
- Cleaning inconsistent labels
- Aggregating key metric
- Generating actionable insights for credit risk strategies

Next steps could include using this cleaned data to build a credit scoring model or visualization dashboard.

Tool Used

- SQL(postgres in DBeaver)
- CSV exports for reporting
- GitHub for project version control

Files

Script-default_credit.sql: Analytical queries for insights

Default_of_credit_card_clients.csv: Row data

Cleaned_default_of_credit_card_clients.csv: All data cleaning queries

Outputs: Folder containing exported CSV result files