Czechoslovakia Banking Financial Data Analysis

Introduction

The Czechoslovakia Bank has provided a dataset containing information about its financial activities for the past 5 years. The dataset consists of the following tables:

1. **Account**: This table contains information about the accounts held by the bank's clients. It includes the account ID, the date the account was opened, the associated client ID, and the account type.
2. **Card**: This table contains information about the card issued by the bank. It includes the card ID, the date the card was issued, and the card type.
3. **Client**: This table contains information about the bank's clients. It includes the client ID, the client's birthdate, gender, and the district where the client lives.
4. **Disposition**: This table contains information about the relationship between clients and their accounts. It includes the disposition ID, the client ID associated with the disposition, and the type of disposition (e.g., owner, authorized person, etc.).
5. **District**: This table contains information about the various districts in Czechoslovakia. It includes the district ID, the name of the district, and various demographic and economic indicators for the district.
6. **Loan**: This table contains information about the loans issued by the bank. It includes the loan ID, the date the loan was issued, the account ID associated with the loan, the amount of the loan.
7. **Order**: This table contains information about the orders issued by the bank's clients. It includes the order ID, the account ID associated with the order, the date the order was issued, and a description of the order.
8. **Transaction**: This table contains information about the transactions made by the bank's clients. It includes the transaction ID, the account ID associated with the transaction, the transaction date, the type of transaction, and the transaction amount.

account.csv

Account

COLUMN NAME	TYPE	DESCRIPTION
# account_id	integer	identification of the account
# district_id	integer	location of the branch
T frequency	string	frequency of issuance of statements: "POPLATEK MESICNE" stands for monthly issuance "POPLATEK TYDNE" stands for weekly issuance "POPLATEK PO OBRATU" stands for issuance after transaction
# date	integer	date of creating of the account: in the form YYMMDD

card.csv

Credit Card

COLUMN NAME	TYPE	DESCRIPTION
# card_id	integer	record identifier
# disp_id	integer	disposition to an account
T type	string	type of card: possible values are "junior", "classic", "gold"
T issued	string	issue date: in the form YYMMDD

client.csv

Client

COLUMN NAME	TYPE	DESCRIPTION
# client_id	integer	client identifier
# birth_number	integer	birthday and sex: the number is in the form YYMMDD for men, the number is in the form YYMM+50DD for women, where YYMMDD is the date of birth
# district_id	integer	address of the client

disp.csv

Disposition

COLUMN NAME	TYPE	DESCRIPTION
# **disp_id**	integer	record identifier
# **client_id**	integer	identification of a client
# **account_id**	integer	identification of an account
T **type**	string	type of disposition (owner/user): only owner can issue permanent orders and ask for a loan

district.csv

Demographic

COLUMN NAME	TYPE	DESCRIPTION
# a1	integer	district code
T a2	string	district name
T a3	string	region
# a4	integer	no. of inhabitants
# a5	integer	no. of municipalities with inhabitants < 499
# a6	integer	no. of municipalities with inhabitants 500-1999
# a7	integer	no. of municipalities with inhabitants 2000-9999
# a8	integer	no. of municipalities with inhabitants >10000
# a9	integer	no. of cities
# a10	decimal	ratio of urban inhabitants
# a11	integer	average salary
# a12	decimal	unemployment rate '95
# a13	decimal	unemployment rate '96
# a14	integer	no. of entrepreneurs per 1000 inhabitants
# a15	integer	no. of committed crimes '95
# a16	integer	no. of committed crimes '96

loan.csv

Loan

COLUMN NAME	TYPE	DESCRIPTION
# loan_id	integer	record identifier
# account_id	integer	identification of the account
# date	integer	date when the loan was granted: in the form YYMMDD
# amount	integer	amount of money
# duration	integer	duration of the loan
# payments	decimal	monthly payments
T status	string	status of paying off the loan: 'A' stands for contract finished, no problems, 'B' stands for contract finished, loan not payed, 'C' stands for running contract, OK so far, 'D' stands for running contract, client in debt

order.csv

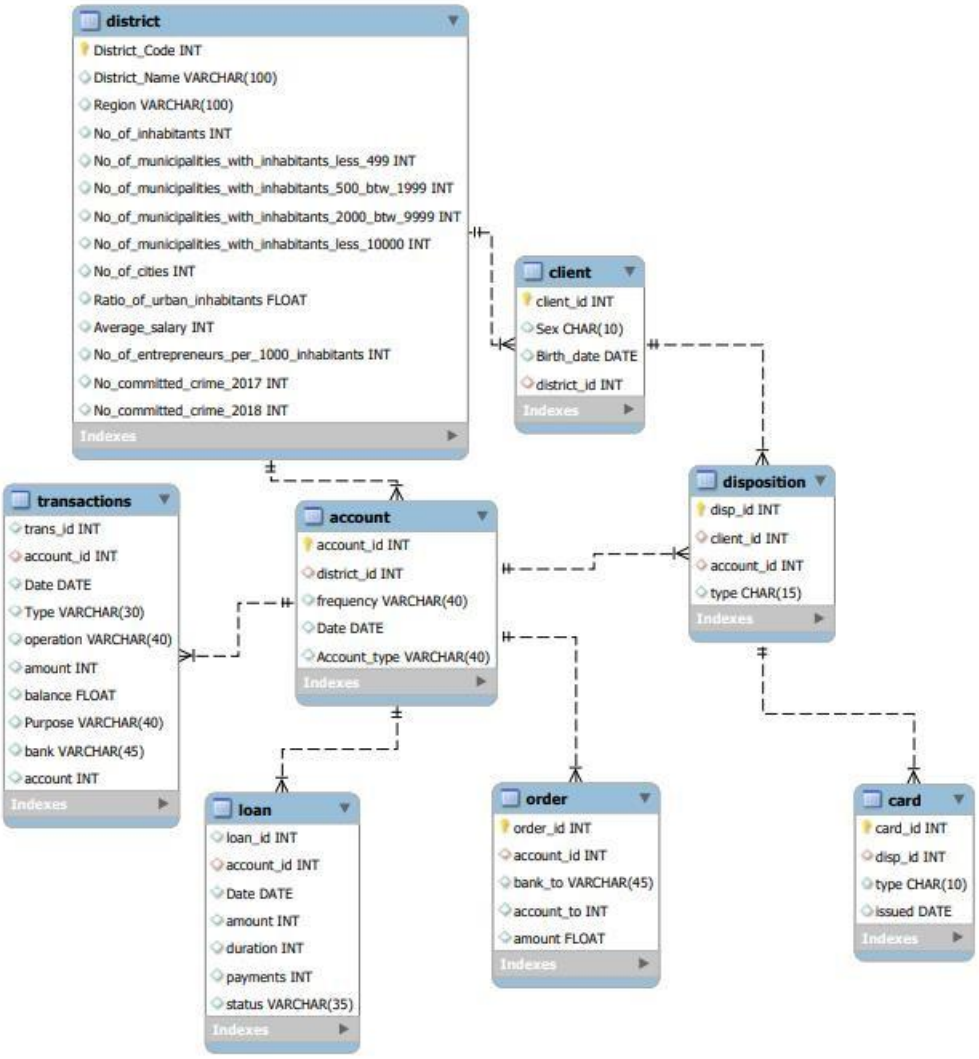
Permanent Order (Debits Only)

COLUMN NAME	TYPE	DESCRIPTION
# order_id	integer	record identifier
# account_id	integer	account, the order is issued for
T bank_to	string	bank of the recipient: each bank has unique two-letter code
# account_to	integer	account of the recipient
# amount	decimal	debited amount
T k_symbol	string	characterization of the payment: "POJISTNE" stands for insurrance payment "SIPO" stands for household payment "LEASING" stands for leasing "UVER" stands for loan payment

Transaction

COLUMN NAME	TYPE	DESCRIPTION
# trans_id	integer	record identifier
# account_id	integer	account the transaction deals with
# date	integer	date of transaction: in the form YYMMDD
T type	string	+/- transaction: "PRIJEM" stands for credit "VYDAJ" stands for withdrawal
T operation	string	mode of transaction: "VYBER KARTOU" credit card withdrawal "VKLAD" credit in cash "PREVOD Z UCTU" collection from another bank "VYBER" withdrawal in cash "PREVOD NA UCET" remittance to another bank
# amount	decimal	amount of money
# balance	decimal	balance after transaction
T k_symbol	string	characterization of the transaction: "POJISTNE" Insurance payment "SLUZBY" payment for statement "UROK" interest credited "SANKC. UROK" sanction Interest If negative balance "SIPO" household "DUCHOD" old-age pension "UVER" loan payment
T bank	string	bank of the partner: each bank has unique two-letter code
# account	integer	account of the partner

ER Diagram


district
District_Code INT
District_Name VARCHAR(100)
Region VARCHAR(100)
No_of_inhabitants INT
No_of_municipalities_with_inhabitants_less_499 INT
No_of_municipalities_with_inhabitants_500_btw_1999 INT
No_of_municipalities_with_inhabitants_2000_btw_9999 INT
No_of_municipalities_with_inhabitants_less_10000 INT
No_of_cities INT
Ratio_of_urban_inhabitants FLOAT
Average_salary INT
No_of_entrepreneurs_per_1000_inhabitants INT
No_committed_crime_2017 INT
No_committed_crime_2018 INT
Indexes
client
client_id INT
Sex CHAR(10)
Birth_date DATE
district_id INT
Indexes
transactions
trans_id INT
account_id INT
Date DATE
Type VARCHAR(30)
operation VARCHAR(40)
amount INT
balance FLOAT
Purpose VARCHAR(40)
bank VARCHAR(45)
account INT
Indexes
account
account_id INT
district_id INT
frequency VARCHAR(40)
Date DATE
Account_type VARCHAR(40)
Indexes
disposition
disp_id INT
client_id INT
account_id INT
type CHAR(15)
Indexes
loan
loan_id INT
account_id INT
Date DATE
amount INT
duration INT
payments INT
status VARCHAR(35)
Indexes
order
order_id INT
account_id INT
bank_to VARCHAR(45)
account_to INT
amount FLOAT
Indexes
card
card_id INT
disp_id INT
type CHAR(10)
issued DATE
Indexes

Ad-hoc Data Analysis

The Czechoslovakia Bank wants to analyse its financial data to gain insights and make informed decisions. The bank needs to identify trends, patterns, and potential risks in its financial operations. They also want to explore the possibility of introducing new financial products or services based on their analysis.

The bank has identified the following questions as important for their analysis:

1. What is the demographic profile of the bank's clients and how does it vary across districts?
2. How the banks have performed over the years. Give their detailed analysis year & month-wise.
3. What are the most common types of accounts and how do they differ in terms of usage and profitability?
4. Which types of cards are most frequently used by the bank's clients and what is the overall profitability of the credit card business?
5. What are the major expenses of the bank and how can they be reduced to improve profitability?

The objective of this analysis is to provide the Czechoslovakia Bank with actionable insights that can help them make informed decisions about their financial operations. The analysis will involve data cleaning, exploratory data analysis, and predictive modelling to identify patterns and trends in the data.