



PKDD'99 Discovery Challenge

Guide to the Financial Data Set

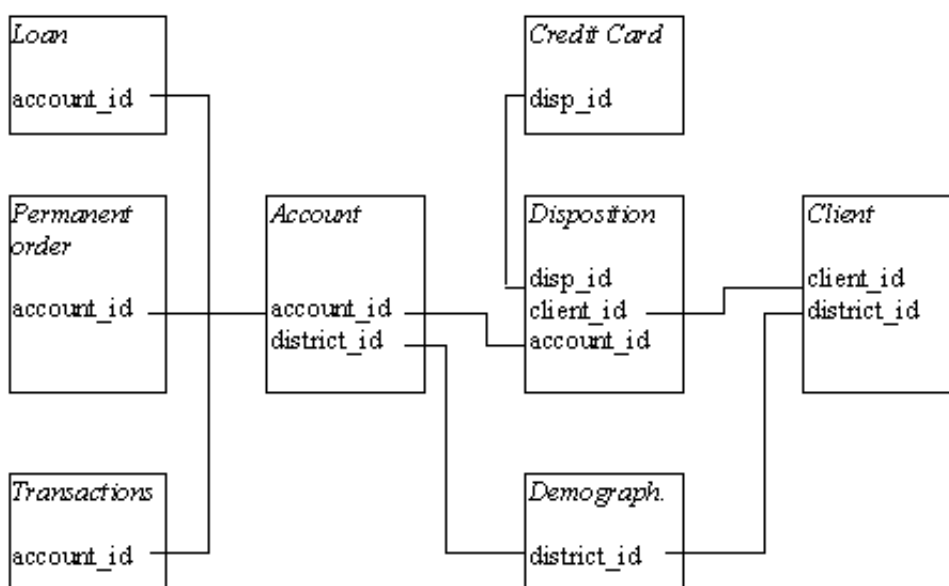
Domain

Once upon a time, there was a bank offering services to private persons. The services include managing of accounts, offering loans, etc.

Task description

The bank wants to improve their services. For instance, the bank managers have only vague idea, who is a good client (whom to offer some additional services) and who is a bad client (whom to watch carefully to minimize the bank losses). Fortunately, the bank stores data about their clients, the accounts (transactions within several months), the loans already granted, the credit cards issued. The bank managers hope to improve their understanding of customers and seek specific actions to improve services. A mere application of a discovery tool will not be convincing for them.

Data description



The data about the clients and their accounts consist of following relations:

- relation **account** (4500 objects in the file ACCOUNT.ASC) - each record describes static characteristics of an account,

- relation **client** (5369 objects in the file CLIENT.ASC) - each record describes characteristics of a client,
- relation **disposition** (5369 objects in the file DISP.ASC) - each record relates together a client with an account i.e. **this relation describes the rights of clients to operate accounts,**
- relation **permanent order** (6471 objects in the file ORDER.ASC) - **each record describes characteristics of a payment order,**
- relation **transaction** (1056320 objects in the file TRANS.ASC) - each record describes one transaction on an account,
- relation **loan** (682 objects in the file LOAN.ASC) - each record describes a loan granted for a given account,
- relation **credit card** (892 objects in the file CARD.ASC) - each record describes a credit card issued to an account,
- relation **demographic data** (77 objects in the file DISTRICT.ASC) - each record describes demographic characteristics of a district.

Each account has both static characteristics (e.g. date of creation, address of the branch) given in relation "account" and dynamic characteristics (e.g. payments debited or credited, balances) given in relations "permanent order" and "transaction". Relation "client" describes characteristics of persons who can manipulate with the accounts. One client can have more accounts, more clients can manipulate with single account; clients and accounts are related together in relation "disposition". Relations "loan" and "credit card" describe some services which the bank offers to its clients; more credit cards can be issued to an account, at most one loan can be granted for an account. Relation "demographic data" gives some publicly available information about the districts (e.g. the unemployment rate); additional information about the clients can be deduced from this.

Relation account

item	meaning	remark
account_id	identification of the account	
district_id	location of the branch	
date	date of creating of the account	in the form YYMMDD
frequency	frequency of issuance of statements (frequência de emissão de extrato)	"POPLATEK MESICNE" stands for monthly issuance (posição mensal de extrato) "POPLATEK TYDNE" stands for weekly issuance (posição semanal de extrato) "POPLATEK PO OBRATU" stands for issuance after transaction (posição depois da transação)

Relation client

item	meaning	remark
client_id	client identifier	

birth number	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	address of the client	

Relation disposition (configuração)

item	meaning	remark
disp_id	record identifier	
client_id	identification of a client	
account_id	identification of an account	
type	type of disposition (owner/user)	only owner can issue permanent orders and ask for a loan

Relation permanent order (debits only)

item	meaning	remark
order_id	record identifier	
account_id	account, the order is issued for	
bank_to	bank of the recipient	each bank has unique two-letter code
account_to	account of the recipient	
amount	debited amount	
K_symbol	characterization of the payment	"POJISTNE" stands for insurance payment (posição para pagamento de seguro) "SIPO" stands for household payment (posição para pagamento de residência) "LEASING" stands for leasing (posição para leasing) "UVER" stands for loan payment (posição para pagamento de empréstimo)

Relation Transaction

item	meaning	remark
trans_id	record identifier	

account_id	account, the transaction deals with	
date	date of transaction	in the form YYMMDD
type	+/- transaction	"PRIJEM" stands for credit (posição para crédito) "VYDAJ" stands for withdrawal (posição para saque)
operation	mode of transaction	"VYBER KARTOU" credit card withdrawal (saque do cartão de crédito) "VKLAD" credit in cash (crédito em dinheiro) "PREVOD Z UCTU" collection from another bank (recebimento de outro banco) "VYBER" withdrawal in cash (saque em dinheiro) "PREVOD NA UCET" remittance to another bank (envio para outro banco)
amount	amount of money	
balance	balance after transaction	
k_symbol	characterization of the transaction	"POJISTNE" stands for insurance payment (posição para pagamento de seguro) "SLUZBY" stands for payment for statement (posição para pagamento de extrato) "UROK" stands for interest credited (posição para juros creditados) "SANKC. UROK" sanction interest if negative balance penalização de juros se conta negativa) "SIPO" stands for household (posição para residência) "DUCHOD" stands for old-age pension (posição para pensão por idade) "UVER" stands for loan payment (posição para pagamento de empréstimo)
bank	bank of the partner	each bank has unique two-letter code
account	account of the partner	

Relation Loan

item	meaning	remark
loan_id	record identifier	
account_id	identification of the account	
date	date when the loan was granted	in the form YYMMDD
amount	amount of money	

duration	duration of the loan	
payments	monthly payments	
status	status of paying off the loan	<p>'A' stands for contract finished, no problems,</p> <p>'B' stands for contract finished, loan not payed,</p> <p>'C' stands for running contract, OK so far,</p> <p>'D' stands for running contract, client in debt</p>

Relation Credit card

item	meaning	remark
card_id	record identifier	
disp_id	disposition to an account	
type	type of card	possible values are "junior", "classic", "gold"
issued	issue date	in the form YYMMDD

Relation Demographic data

item	meaning	remark
A1 = district_id	district code	
A2	district name	
A3	region	
A4	no. of inhabitants	
A5	no. of municipalities with inhabitants < 499	
A6	no. of municipalities with inhabitants 500-1999	
A7	no. of municipalities with inhabitants 2000-9999	
A8	no. of municipalities with inhabitants >10000	
A9	no. of cities	
A10	ratio of urban inhabitants	
A11	average salary	
A12	unemployment rate '95	
A13	unemployment rate '96	
A14	no. of entrepreneurs per 1000 inhabitants	

A15	no. of committed crimes '95	
A16	no. of committed crimes '96	

This database was prepared by Petr Berka and Marta Sochorova.

For possible questions on the data and task description contact Petr Berka. All questions and answers will be published as appendixes to this document.

Asked Questions