

Banking Project Brief

A retail bank's services include managing of accounts, offering loans and credit cards. The objective of this project is to find opportunities to improve the bank's services by finding interesting groups of customers (e.g. to differentiate between who may be good and bad customers) or interesting analysis of customer behaviour (e.g. actual or likely spending patterns).

The sort of info we are looking for is anything that identifies customers who may be valuable and profitable versus customers who may be a bad risk.

To support this analysis the bank stores data about their customers, account transactions, the loans already granted and the credit cards issued. They also have access to various pieces of demographic data about populations in the places where their customers live. By analysing this data we want you to provide any useful insights – **backed up with real data as evidence!** – into customer behaviour.

Download and import the data from this link:

<https://tinyurl.com/kbankproject>

The data about the customers and their accounts consist of following tables:

- **account** (4500 rows in the file ACCOUNT.ASC) - each row describes static characteristics of an account
- **client** (5369 rows in the file CLIENT.ASC) - each row describes a customer
- **disposition** (5369 rows in the file DISP.ASC) - each row relates a customer to account(s) which they hold
- **permanent order** (6471 rows in the file ORDER.ASC) - each row describes characteristics of a payment order
- **transaction** (1056320 rows in the file TRANS.ASC) - each row describes one transaction on an account
- **loan** (682 rows in the file LOAN.ASC) - each row describes a loan granted for an account
- **credit card** (892 rows in the file CARD.ASC) - each row describes a credit card issued to an account
- **demographic data** (77 rows in the file DISTRICT.ASC) - each row describes demographic characteristics of a district

Each account has both static characteristics (e.g. date of creation, address of the branch) given in the “account” table and dynamic characteristics (e.g. payments debited or credited, balances) given in the tables “permanent order” and

“transaction”. The “client” table describes characteristics of the customers using the accounts. It is possible for two customers to have a joint account together; customers and accounts are related by the “disposition” table.

The “loan” and “credit card” tables show the loans made and cards issued to customers. Several credit cards can be issued per account. At most, one loan can be granted for an account.

The table “demographic data” gives some publicly available information about the districts where customers live (e.g. the unemployment rate). You can use this to derive any useful additional analysis of customers.

At the end of the project you need to present back to the group:

1. Summary of the project brief and why your analysis will be an enabler for well-informed decisions for the bank. (10%)
2. The general approach you took, your thought process and decisions. (20%)
3. Working examples of your code, any interesting technical problems solved, lessons learned or new features used. (35%)
4. Outputs: Presentation of any useful data insights, reports and graphs, including underlying numbers (35%)

Table Structures

Table: 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 YYYYMMDD
frequency	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

Table: client

item	meaning	remark
client_id	row identifier	
birth number	identification of client	the number is in the form YYYYMMDD for men, the number is in the form YYYYMM+50DD for women, where YYYYMMDD is the date of birth
district_id	address of the client	

Table: disposition

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

Table: permanent order

item	meaning	remark
order_id	row identifier	
account_id	account, the order is issued for	
bank_to	bank of the recipient	each bank has a unique two-letter code
account_to	account of the recipient	
amount	debited amount	
K_symbol	characterization of the payment	"POJISTNE" stands for insurance payment "SIPO" stands for household "LEASING" stands for leasing "UVER" stands for loan payment

Table: Transaction

item	meaning	remark
trans_id	row identifier	
account_id	account, the transaction deals with	
date	date of transaction	in the form YYMMDD
type	+/- transaction	"PRIJEM" stands for credit "VYDAJ" stands for withdrawal "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
operation	mode of transaction	
amount	amount of money	
balance	balance after transaction	
k_symbol	characterization of the transaction	"POJISTNE" stands for insurance payment "SLUZBY" stands for payment for statement "UROK" stands for interest credited "SANKC. UROK" sanction interest if negative balance "SIPO" stands for household "DUCHOD" stands for old-age pension "UVER" stands for loan payment
bank	bank of the partner	each bank has a unique two-letter code
account	account of the partner	

Table: Loan

item	Meaning	remark
loan_id	row 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	'A' stands for contract finished, no problems, 'B' stands for contract finished, loan not paid, 'C' stands for running contract, OK so far, 'D' stands for running contract, client in debt

Table: Credit card

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

Table: 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	

Example Reporting Ideas

