

The Berka dataset analysis

Project Info

Data source:

<https://www.kaggle.com/datasets/marceloventura/the-berka-dataset?resource=download>

Dataset documentation:

<https://webpages.charlotte.edu/mirsad/itcs6265/group1/index.html>

Project Goal:

From data set documentation:

The Berka dataset, from the 1999 PKDD Discovery Challenge, provides information on the clients, accounts, and transactions of a Czech bank. The original task description of the Discovery Challenge states:

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 loses). 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.

In keeping with the original task description, our project goal is to mine and analyze this bank data in order to extrapolate from it the type of customer who makes a good candidate for a credit card.

Main tasks:

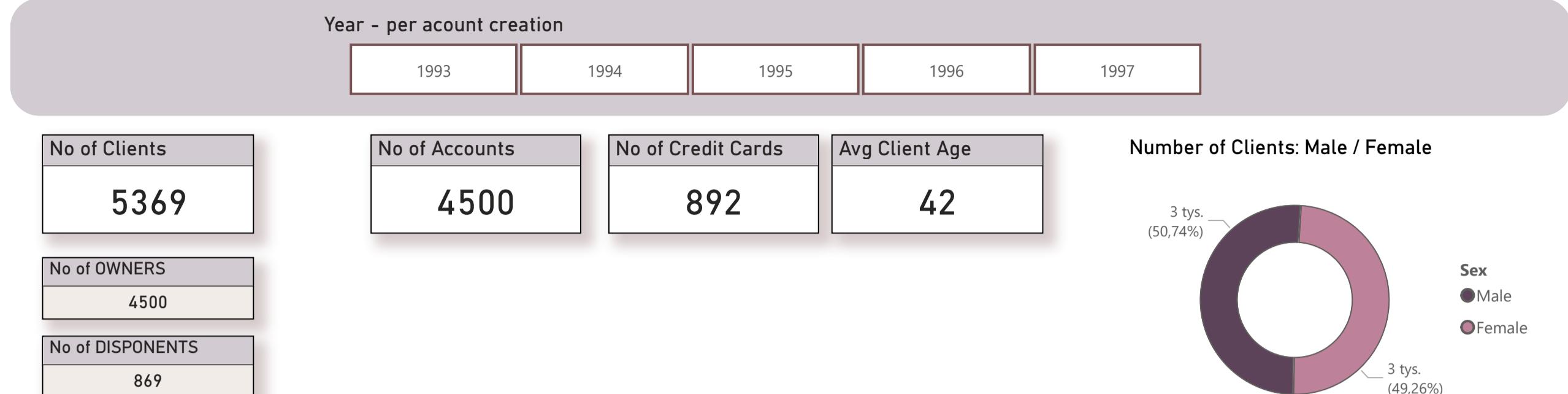
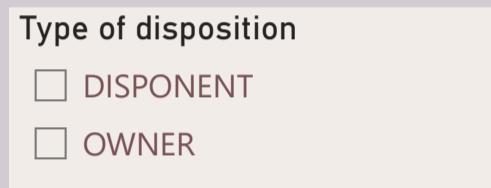
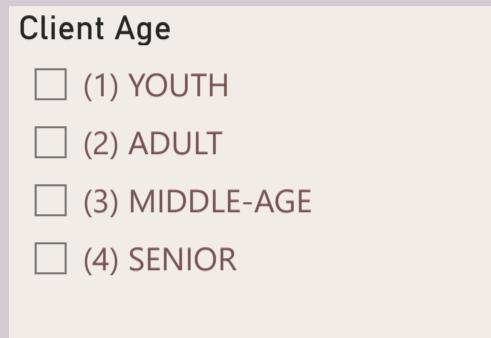
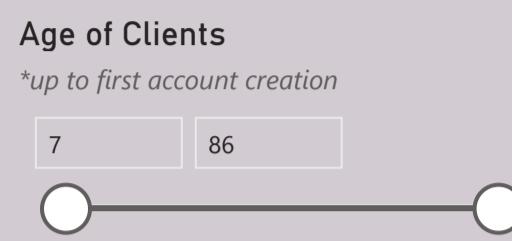
- ◆ load data (ETL - Extract, Transform, Load)
- ◆ create relationships / prepare data model
- ◆ clean data
- ◆ prepare an analysis of customer transactions e.g.: receipts on accounts, payments from accounts, customers making the most transactions, the fewest transactions, regions with the most active customers, types of cards used by customers, loans and other products held by customers etc.

Dashboard Structure:

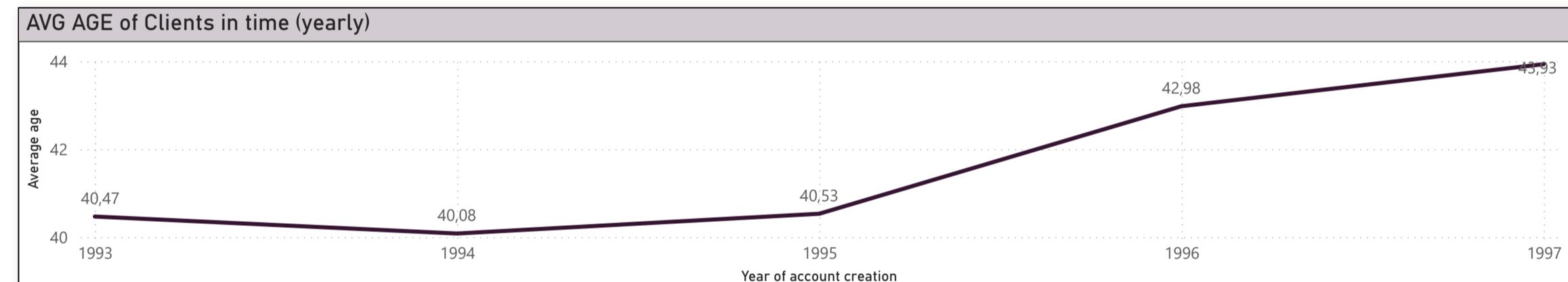
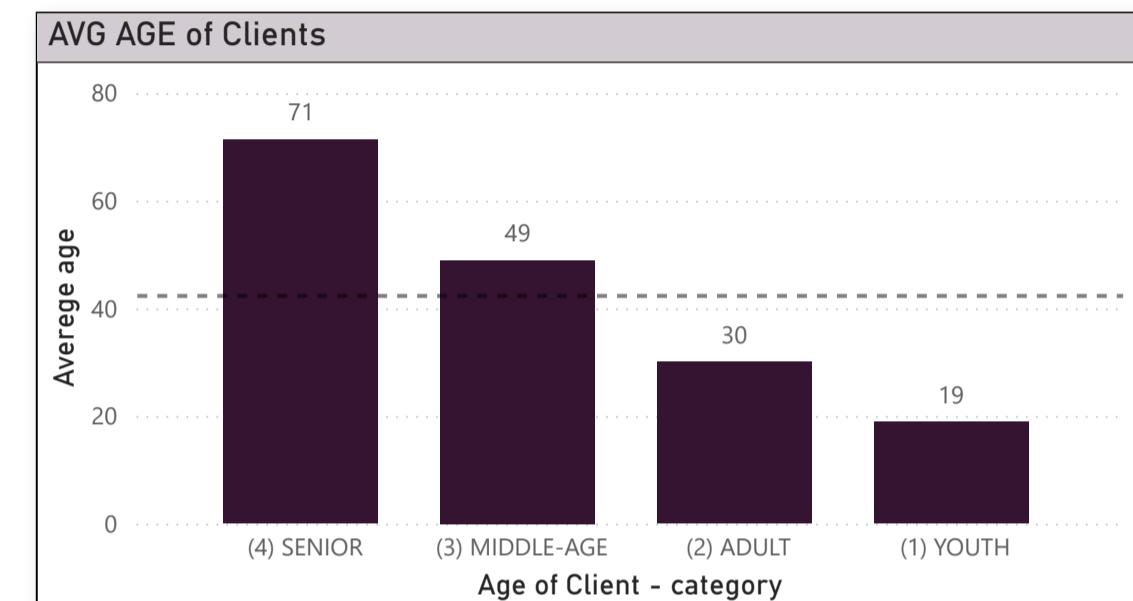
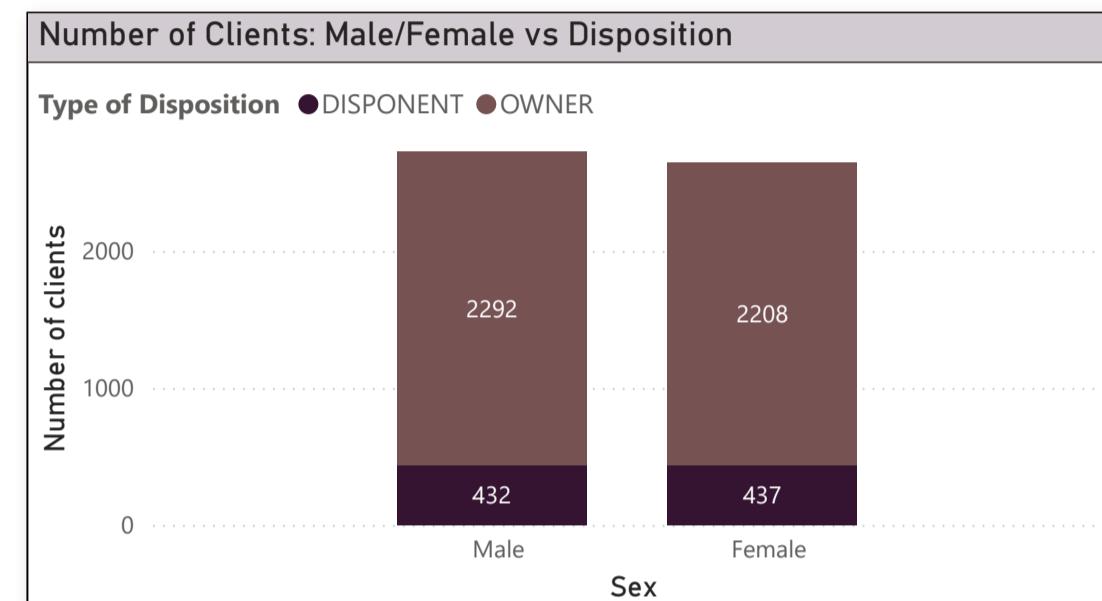
- (i) Info - Data source and project goal
- (ii) Clients
- (iii) Accounts
- (iv) Transactions
- (v) Loans
- (vi) Orders

The Berka dataset analysis

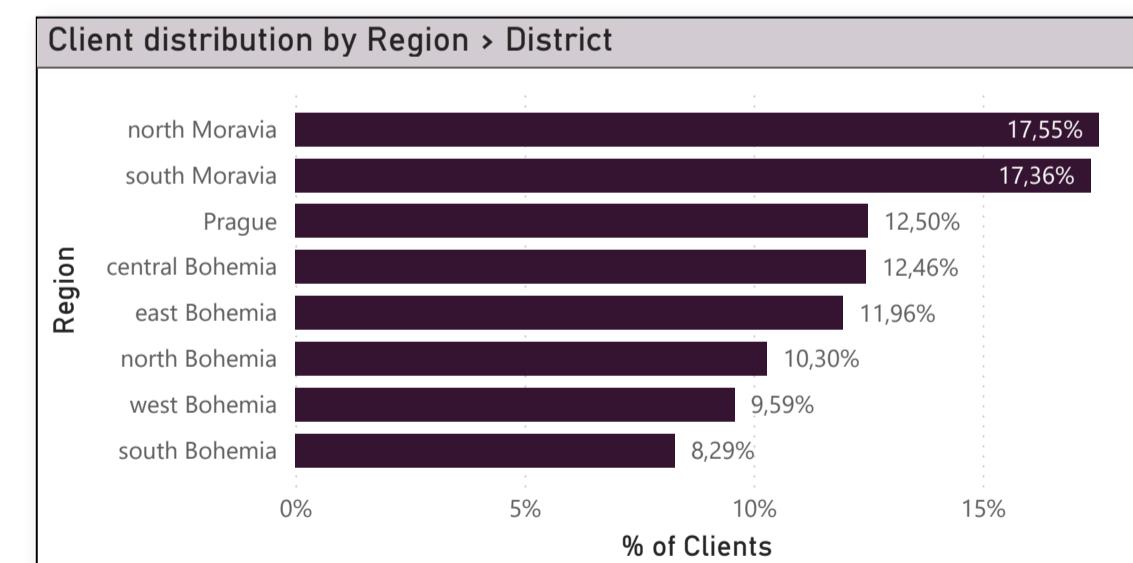
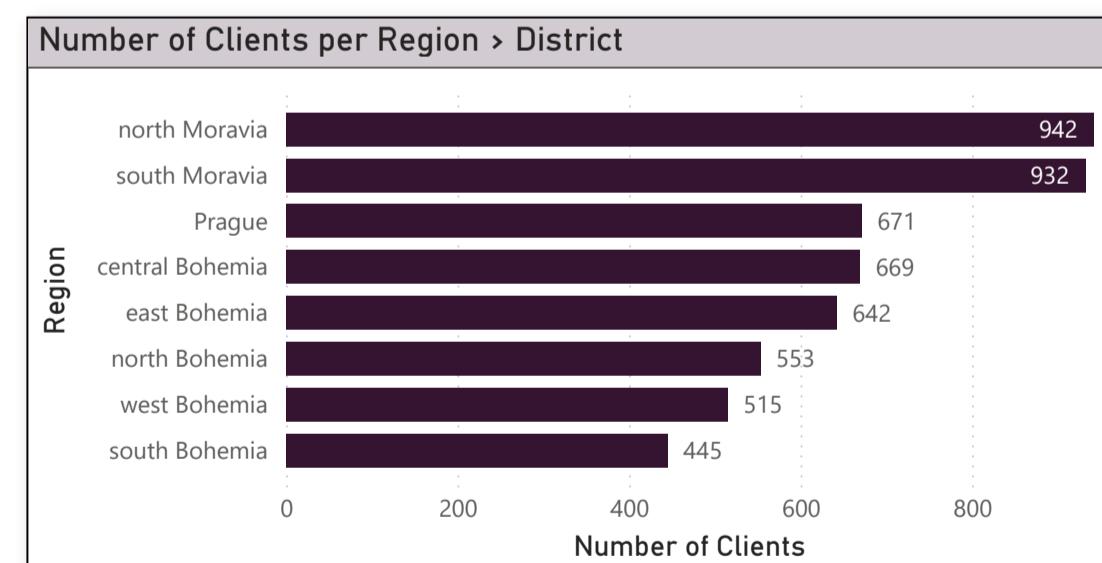
Bank Clients analysis

[Reset Filters](#)


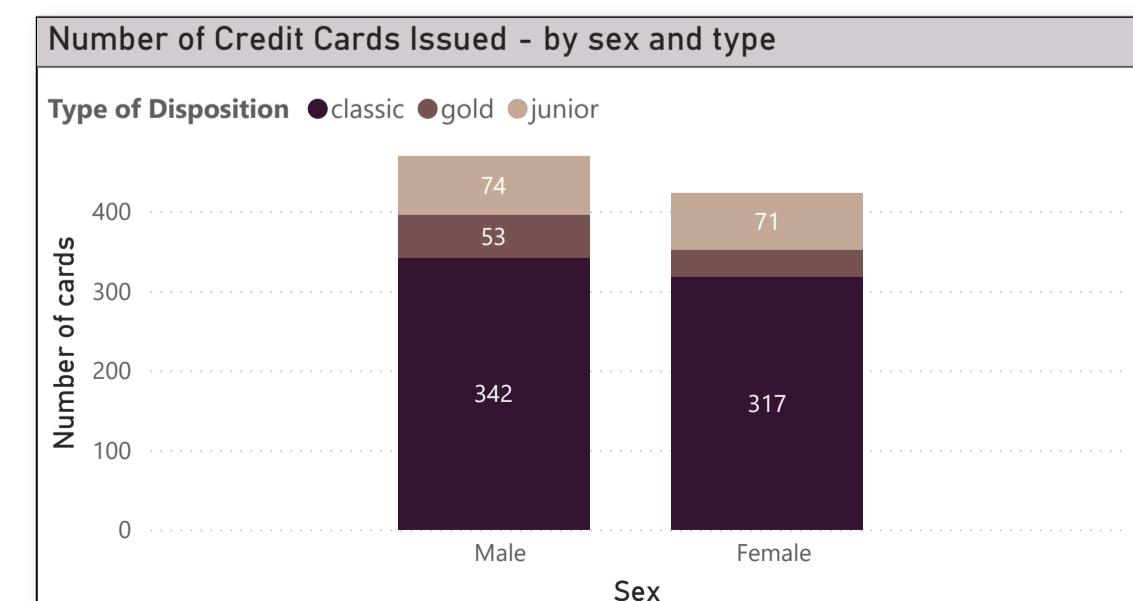
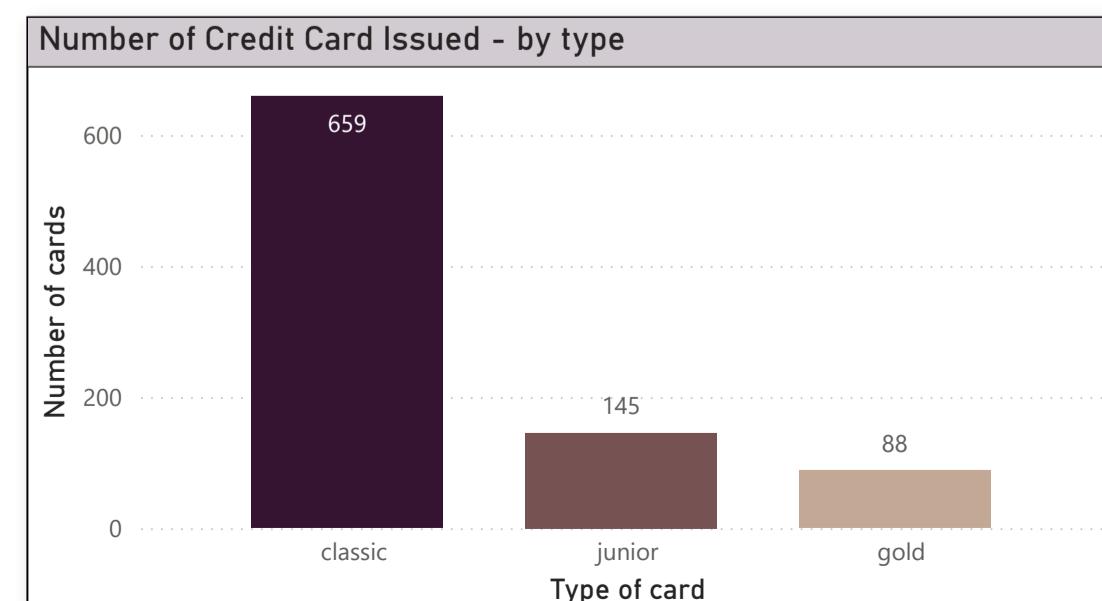
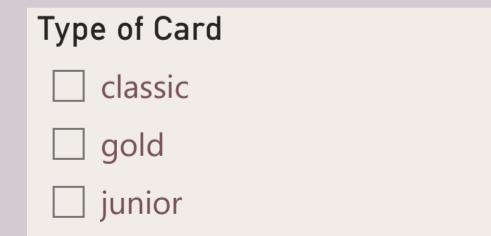
CLIENTS INFO



DEMOGRAPHICS



CREDIT CARDS



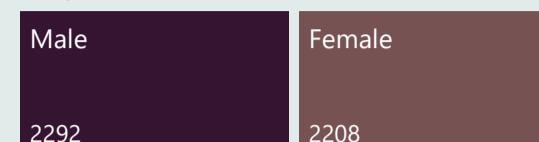
The Berka dataset analysis

Accounts analysis

[Reset Filters](#)


Accounts - per Owner

*Disponents not included



Accounts - per Users

*One account can have Owner and Disponent



Account Status

*Inactive - account not used more than 60 days



Frequency of statement issuance

- ISSUANCE AFTER TRANSACTION (TI)
- MONTHLY ISSUANCE (MI)
- WEEKLY ISSUANCE (WI)

Type of disposition

- DISPONENT
- OWNER

Client Age

- (1) YOUTH
- (2) ADULT
- (3) MIDDLE-AGE
- (4) SENIOR

Account_id

- Wyszukaj
- 1
- 10
- 100
- 1000

Year - per account creation

1993	1994	1995	1996	1997
------	------	------	------	------

No of Accounts

4500

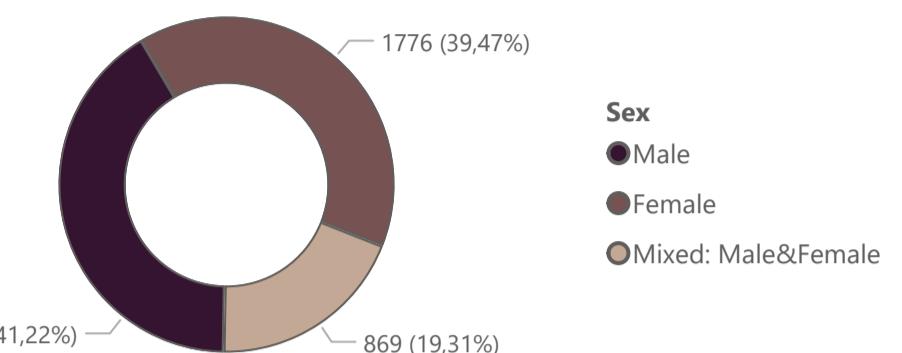
No of Clients

5369

No of Credit Cards

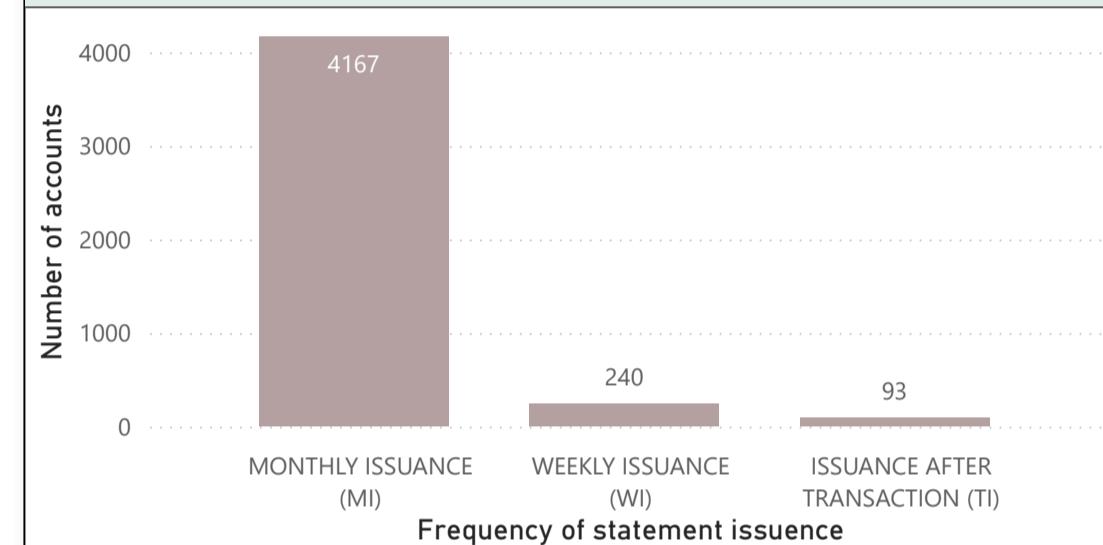
892

Number of Accounts - per Users (includes Owners and Disponents)

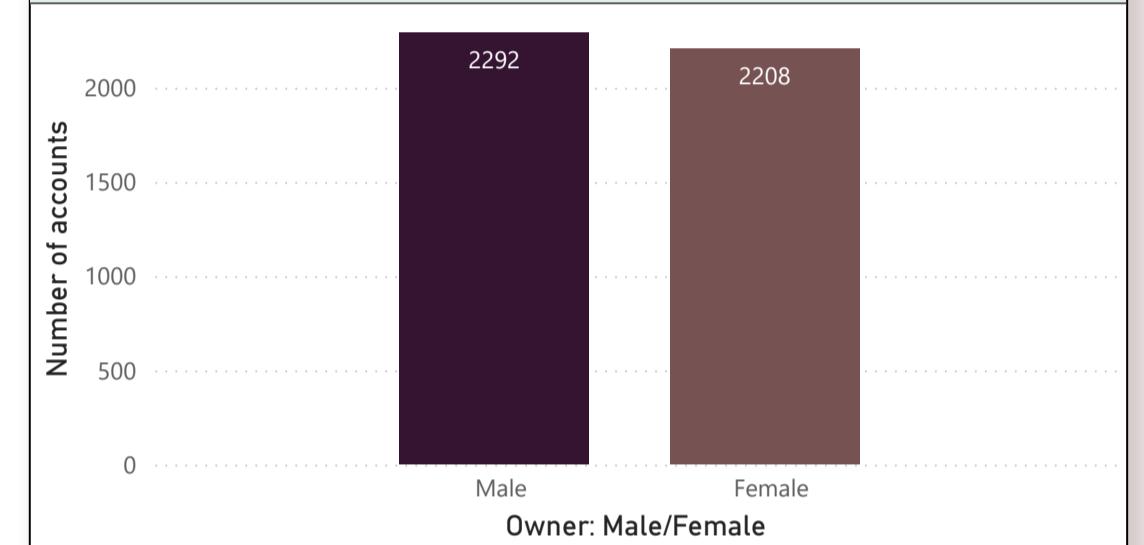


ACCOUNTS INFO

Number of Accounts: per frequency of statement issuance

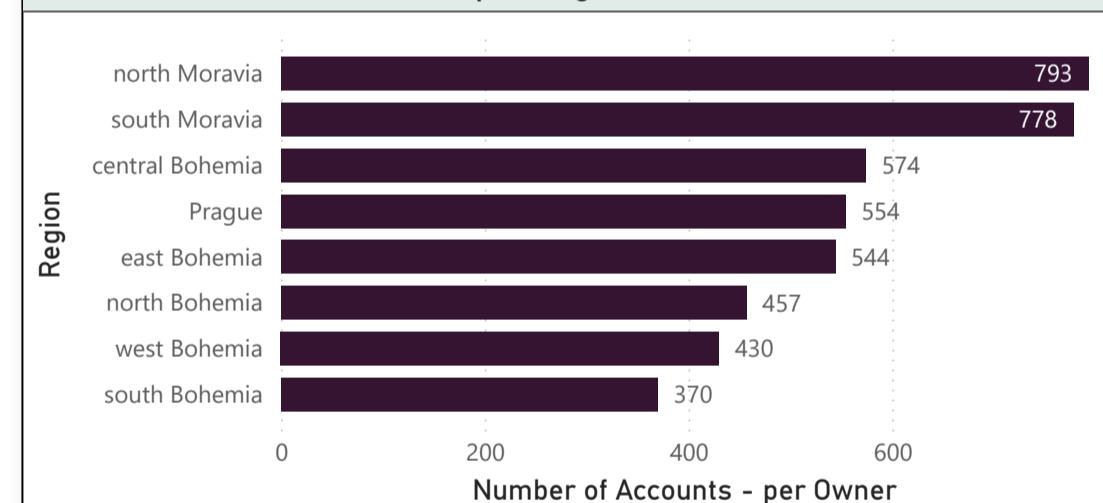


Number of Accounts: per Owner: Male/Female

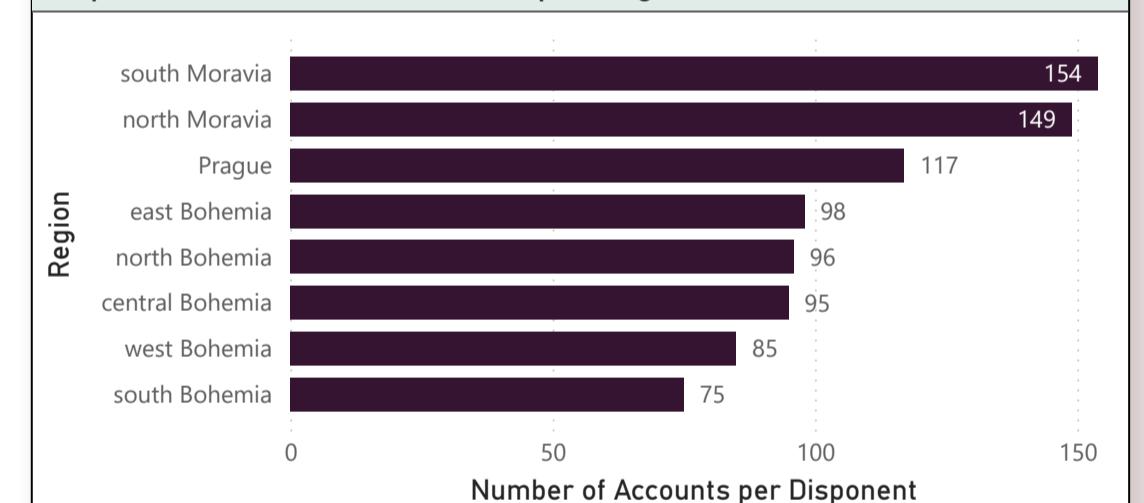


DEMOGRAPHICS

Owners: Number of Accounts per Region > District



Dispontents: Number of Accounts per Region > District



ACCOUNTS DETAILS

Account ID selected

No/or multiple selection

Bank: Date of last trans.

1998-12-31

Account: Date of last trans.

(Puste)

Days since last trans.

858

Avg no of trans.

235

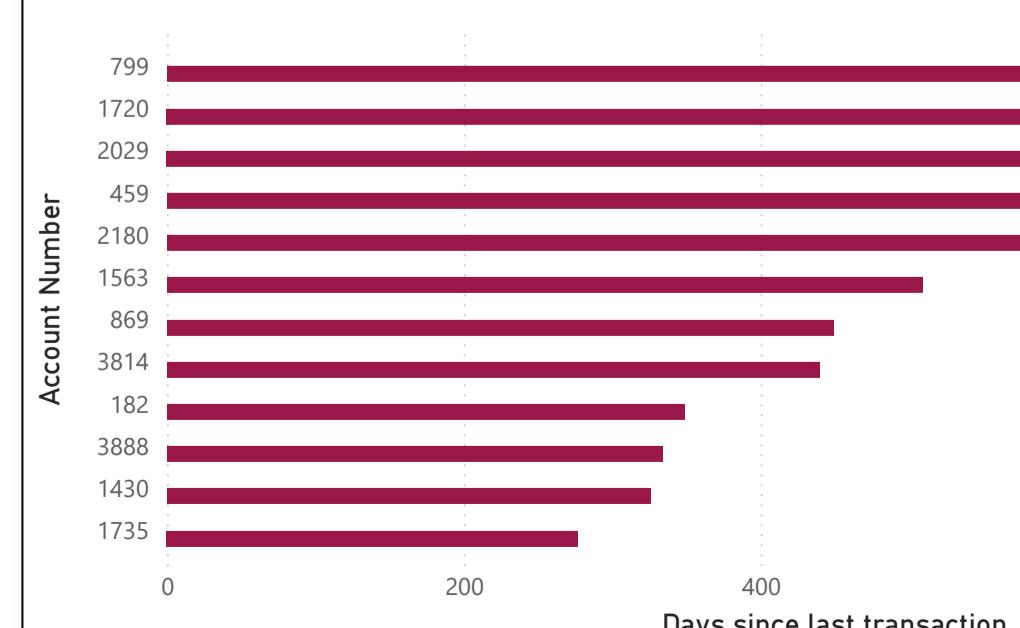
No of trans.

1 mln

Account Numbers: Time (days) since last transaction

*Inactive - account not used more than 60 days

Status ● Active ● Inactive



Accounts: Dispontent vs Owner

Account No DISPONENT OWNER

998	1	1
997		1
996		1
995		1
994		1
993		1
9928	1	1
992		1
991		1
990		1
99		1
9897		1
989		1
9883		1
988		1
987	1	1
Suma	869	4500

The Berka dataset analysis

[Reset Filters](#)

Transactions Analysis



Type of disposition

- DISPONENT
- OWNER

Type of Transaction

- CREDIT
- DEBIT (WITHDRAWAL)
- WITHDRAWAL IN CASH

Operation / Mode of Transaction

- COLLECTION FROM ANOTHER BANK
- CREDIT CARD WITHDRAWAL
- CREDIT IN CASH
- NULL
- REMITTANCE TO ANOTHER BANK
- WITHDRAWAL IN CASH

Character of Transaction

Wszystkie ▼

Year

1993 1994 1995 1996 1997 1998

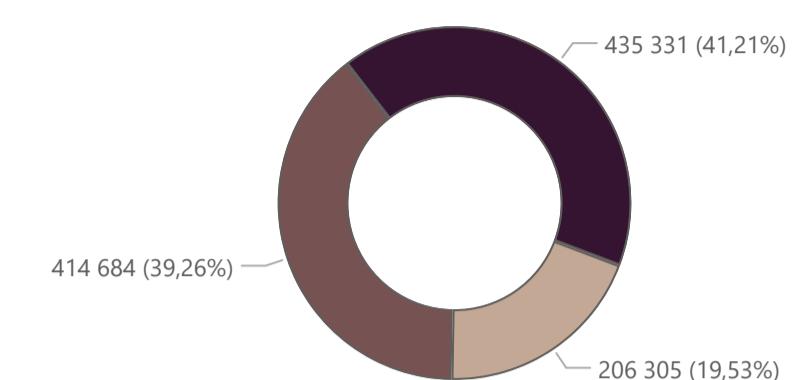
Transaction Date

1993-01-01 1998-12-31

No of Transactions	Total Deposit (Credit)	No of Accounts	No of Clients
1 056 320	3 227 472 425	4500	5369
Avg no of trans. per Account	Total Withdrawals (Debit)	No of OWNERS	No of DISPONENTS
235	-3 030 321 136	4500	869
Avg no of trans. per Client			
235			

Number of Transactions - per Users (includes Owners and Disponents)

Sex ● Female ● Male ○ Mixed: Male&Female



TRANSACTIONS INFO

Total Split per Years

Number and amount of transactions - Total

Type of Transaction	Operation / Mode of Transaction	Character of Transaction	Number of Transactions	Total Deposits	Total Withdrawals
DEBIT (WITHDRAWAL)	REMITTANCE TO ANOTHER BANK	HOUSEHOLD PAYMENT	115 254		-476 103 616
DEBIT (WITHDRAWAL)	WITHDRAWAL IN CASH	HOUSEHOLD PAYMENT	2 811		-22 402 800
DEBIT (WITHDRAWAL)	REMITTANCE TO ANOTHER BANK	INSURANCE PAYMENT	18 477		-24 151 293
DEBIT (WITHDRAWAL)	WITHDRAWAL IN CASH	INSURANCE PAYMENT	23		-23 900
CREDIT	NULL	INTEREST CREDITED	183 114	27 470 690	
DEBIT (WITHDRAWAL)	REMITTANCE TO ANOTHER BANK	LOAN PAYMENT	13 580		-55 252 303
CREDIT	COLLECTION FROM ANOTHER BANK	NULL	34 888	614 007 835	
CREDIT	CREDIT IN CASH	NULL	156 743	2 418 521 782	
DEBIT (WITHDRAWAL)	CREDIT CARD WITHDRAWAL	NULL	8 036		-18 170 400
DEBIT (WITHDRAWAL)	REMITTANCE TO ANOTHER BANK	NULL	60 972		-117 130 574
DEBIT (WITHDRAWAL)	WITHDRAWAL IN CASH	NULL	258 009		-2 105 758 410
WITHDRAWAL IN CASH	WITHDRAWAL IN CASH	NULL	16 666		-208 603 758
CREDIT	COLLECTION FROM ANOTHER BANK	OLD-AGE PENSION PAYMENT	30 338	167 472 118	
DEBIT (WITHDRAWAL)	WITHDRAWAL IN CASH	PAYMENT OF STATEMENT	155 832		-2 686 418
DEBIT (WITHDRAWAL)	WITHDRAWAL IN CASH	SANCTION INTEREST IF NEGATIVE BALANCE	1 577		-37 663
Suma			1 056 320	3 227 472 425	-3 030 321 136

Transaction ID

○ Wyszukaj
 1
 10
 10000
 100008
 100009
 10001
 100010
 100011

Number of Transactions wg Year



TRANSACTIONS DETAILS

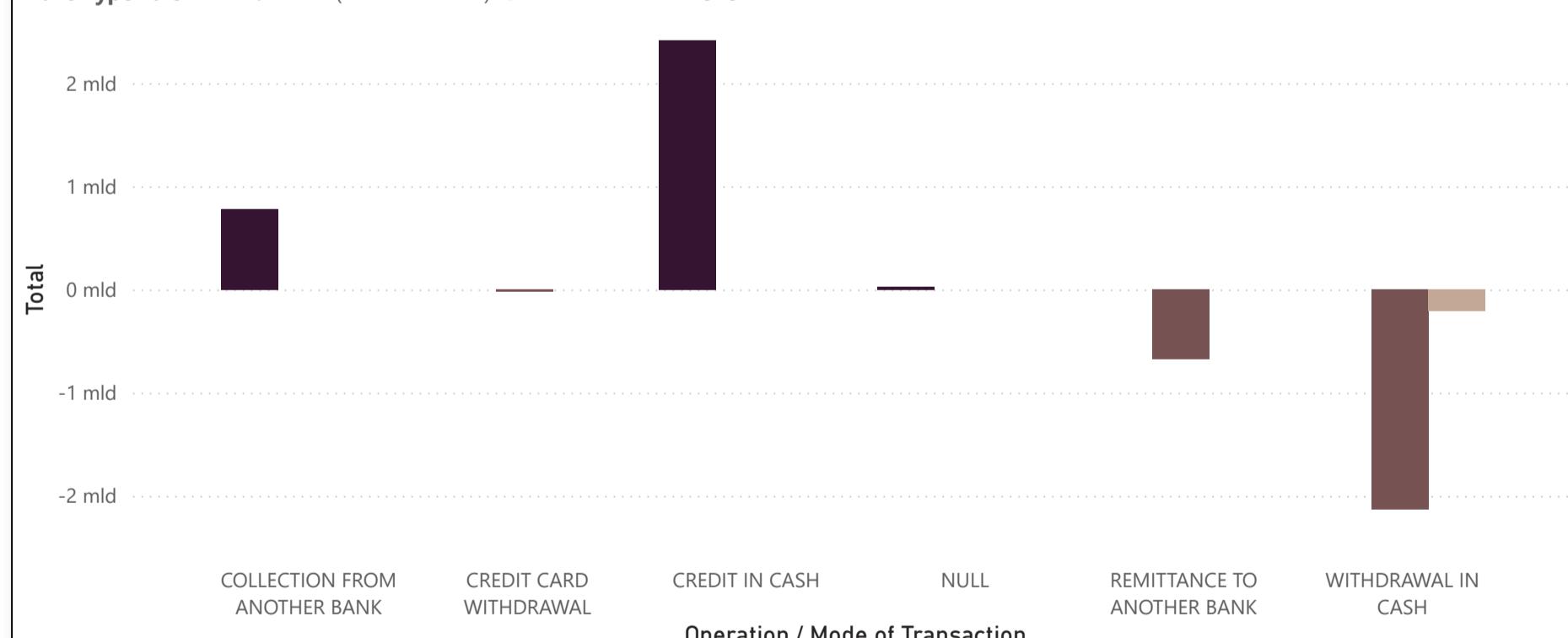
Client Age

(1) YOUTH
 (2) ADULT
 (3) MIDDLE-AGE
 (4) SENIOR

- (1) Transactions Number
- (2) Transactions Amount
- (3) Avg Trans Number per Client
- (4) Avg Trans Amount per Client

Selected Metrics per Type and Mode of Transaction

Trans Type ● CREDIT ● DEBIT (WITHDRAWAL) ● WITHDRAWAL IN CASH



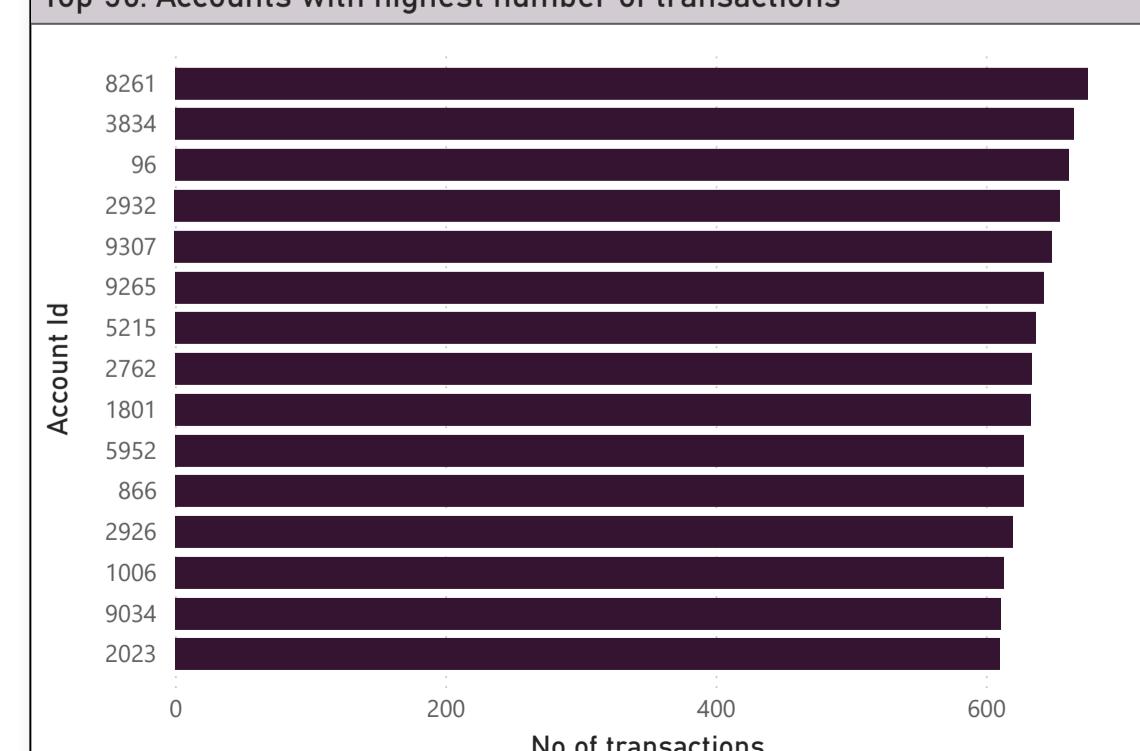
Account Id

○ Wyszukaj
 1
 10
 100
 1000

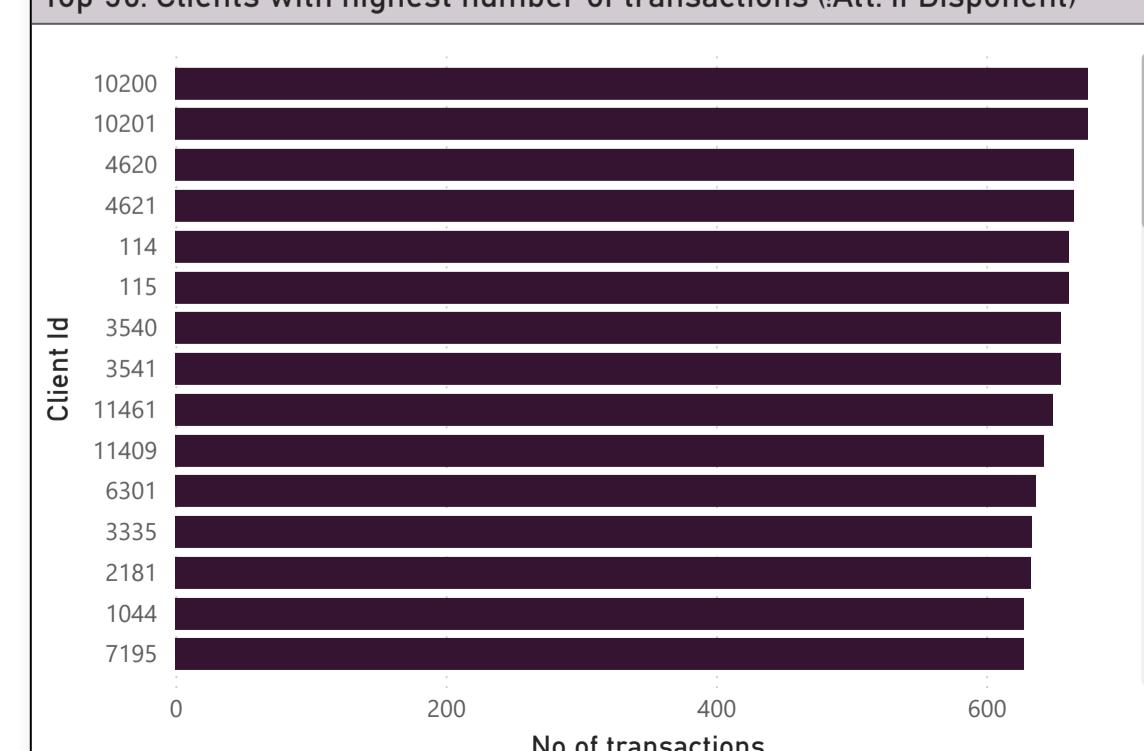
Client Id

○ Wyszukaj
 1
 10
 100
 1000

Top 50: Accounts with highest number of transactions



Top 50: Clients with highest number of transactions (!Att. if Disponent)



The Berka dataset analysis

[Reset Filters](#)

Loans Analysis

Loans - per Sex assigned to Account

*One account can have Owner and Disponent

Female	Male	Mixed: ...
272	265	145

Loan Status

Wszystkie ▼

Loan Duration

Wszystkie ▼

Loan Contract Status

- CONTRACT FINISHED
- RUNNING CONTRACT

Loan Payoff Status

- DELAYS
- NOT PAID
- PAYMENTS ON TIME

Type of disposition

- DISPONENT
- OWNER

Client Age

- (1) YOUTH
- (2) ADULT
- (3) MIDDLE-AGE
- (4) SENIOR

Year

1993	1994	1995	1996	1997	1998
------	------	------	------	------	------

Number of Loans

682

Avg Loan Duration

36

AVG Duration in MONTHS

Total Amount of Loans

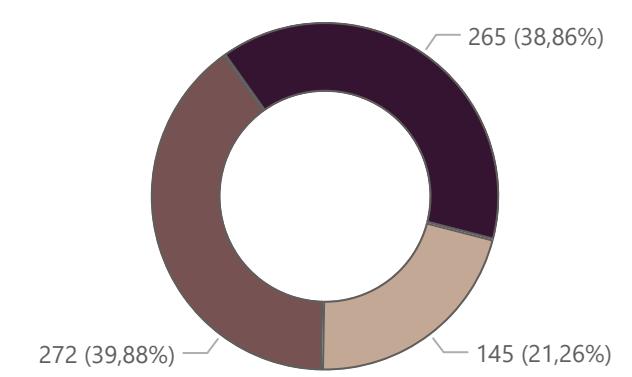
103 261 740

Amount of Remining Instalments RUNNING CONTRACT ONLY

11 115

Number Loans - per Users (includes Owners and Disponents)

Sex ● Female ● Male ● Mixed: Male&Female



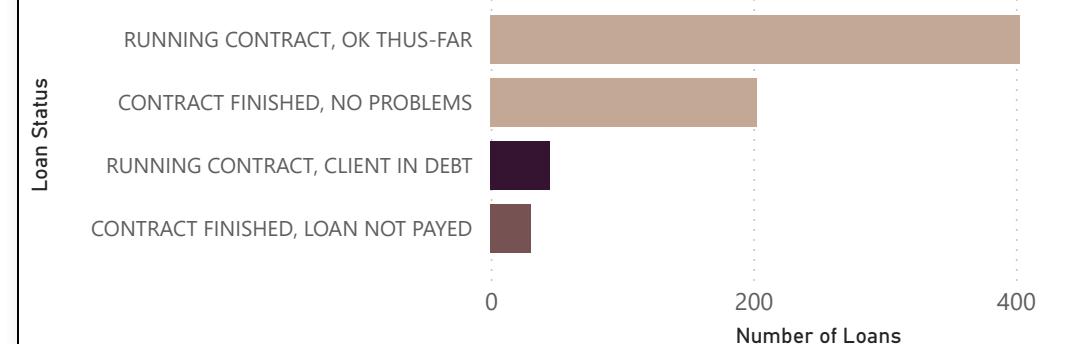
LOANS INFO

Total Number and Amount of Loans per status

Loan Contract Status	Loan PayOff Status	Number of Loans	Total Amount of Loans
CONTRACT FINISHED	NOT PAID	31	4 362 348
CONTRACT FINISHED	PAYMENTS ON TIME	203	18 603 216
RUNNING CONTRACT	DELAYS	45	11 217 804
RUNNING CONTRACT	PAYMENTS ON TIME	403	69 078 372
Suma		682	103 261 740

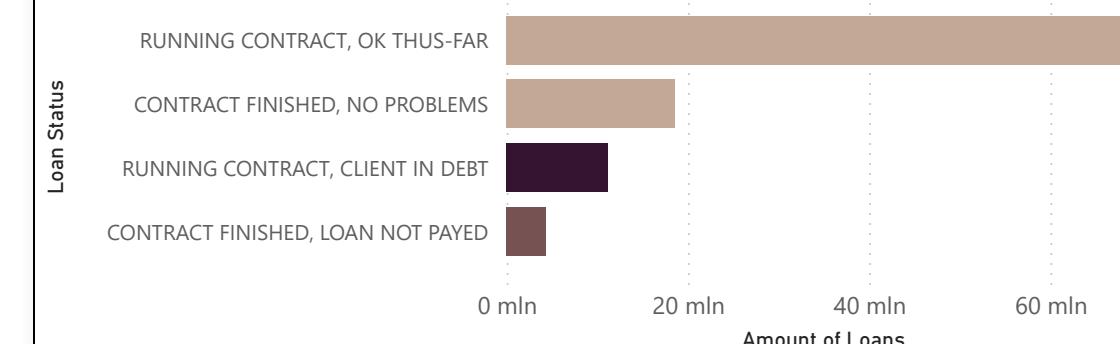
Total Number of Loans per status

Payoff Status ● DELAYS ● NOT PAID ● PAYMENTS ON TIME



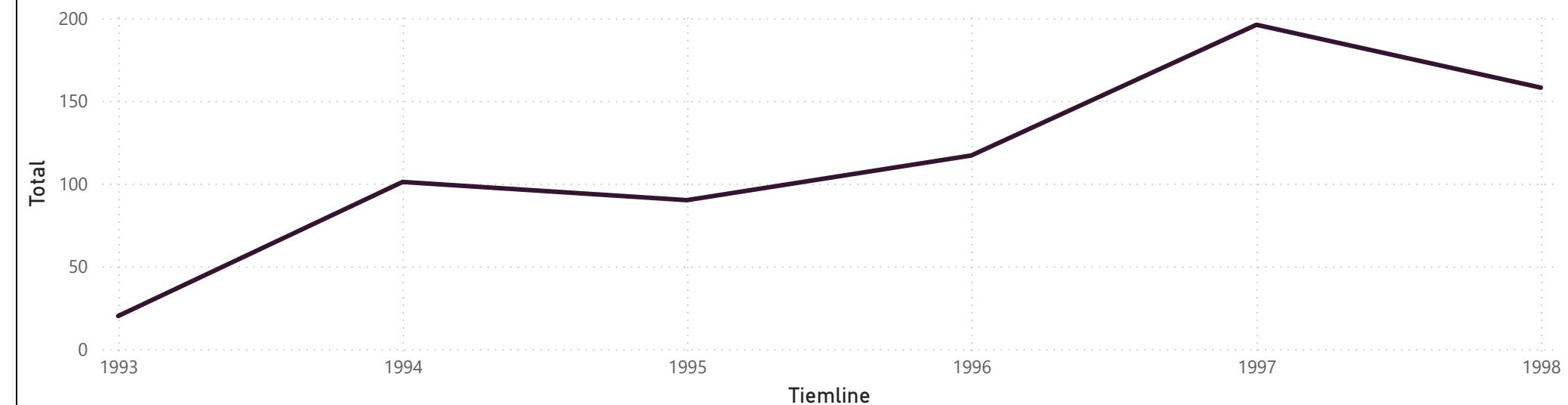
Total Amount of Loans per status

Payoff Status ● DELAYS ● NOT PAID ● PAYMENTS ON TIME



- Amount of Loans
- Number of Loans

Selected Metrics (Amount or Numer of Loans) in Time



LOANS DETAILS

Loan ID selected

No/or multiple selection

Remining Instalments per Loan ID (MONTHS)

(Puste)

Remining Instalments per Loan ID (AMOUNT)

(Puste)

Loan ID

Wyszukaj

4959

4961

4962

4967

4968

4973

4986

5088

Loans details

Loan ID	Contract Status	PayOff Status	Loan Amount	Duration (months)	Loan Granted	Bank Last Transaction	Remining Instalments (months)	Remaning Instalments (AMOUNT)	Info on Remining Instalments
4959	CONTRACT FINISHED	● PAYMENTS ON TIME	80 952	24	1994-01-05	1998-12-31	0	0	<- INFO PROVIDED
4961	CONTRACT FINISHED	◆ NOT PAID	30 276	12	1996-04-29	1998-12-31	0	0	✖ NO INFO
4962	CONTRACT FINISHED	● PAYMENTS ON TIME	30 276	12	1997-12-08	1998-12-31	0	0	<- INFO PROVIDED
4973	CONTRACT FINISHED	● PAYMENTS ON TIME	165 960	24	1996-05-02	1998-12-31	0	0	<- INFO PROVIDED
4986	CONTRACT FINISHED	● PAYMENTS ON TIME	102 876	12	1997-08-10	1998-12-31	0	0	<- INFO PROVIDED
4996	CONTRACT FINISHED	● PAYMENTS ON TIME	88 440	12	1996-11-06	1998-12-31	0	0	<- INFO PROVIDED
5002	CONTRACT FINISHED	● PAYMENTS ON TIME	104 808	12	1994-05-31	1998-12-31	0	0	<- INFO PROVIDED
5005	CONTRACT FINISHED	● PAYMENTS ON TIME	27 456	12	1997-04-10	1998-12-31	0	0	<- INFO PROVIDED
5032	CONTRACT FINISHED	● PAYMENTS ON TIME	123 696	48	1994-11-03	1998-12-31	0	0	<- INFO PROVIDED
5045	CONTRACT FINISHED	◆ NOT PAID	187 224	24	1995-05-01	1998-12-31	0	0	✖ NO INFO
5088	CONTRACT FINISHED	● PAYMENTS ON TIME	91 152	24	1996-08-22	1998-12-31	0	0	<- INFO PROVIDED
5103	CONTRACT FINISHED	● PAYMENTS ON TIME	12 540	12	1997-11-28	1998-12-31	0	0	<- INFO PROVIDED
5117	CONTRACT FINISHED	● PAYMENTS ON TIME	76 944	12	1997-08-20	1998-12-31	0	0	<- INFO PROVIDED
5126	CONTRACT FINISHED	◆ NOT PAID	208 128	48	1994-07-24	1998-12-31	0	0	✖ NO INFO
5130	CONTRACT FINISHED	● PAYMENTS ON TIME	24 312	12	1994-05-11	1998-12-31	0	0	<- INFO PROVIDED
5131	CONTRACT FINISHED	● PAYMENTS ON TIME	48 624	24	1995-01-08	1998-12-31	0	0	<- INFO PROVIDED



The Berka dataset analysis

Orders Analysis

[Reset Filters](#)


Orders - per Sex assigned to Account

*One account can have Owner and Disponent

Male	Female	Mixed: ...
2559	2515	1397

Characterization of the Payment

- HOUSEHOLD PAYMENT
- INSURANCE PAYMENT
- LEASING PAYMENT
- LOAN PAYMENT
- NO INFO

Type of disposition

- DISPONENT
- OWNER

Client Age

- (1) YOUTH
- (2) ADULT
- (3) MIDDLE-AGE
- (4) SENIOR

Year

1993	1994	1995	1996	1997	1998
------	------	------	------	------	------

No of Orders

6471

Sum of Orders

21 228 994

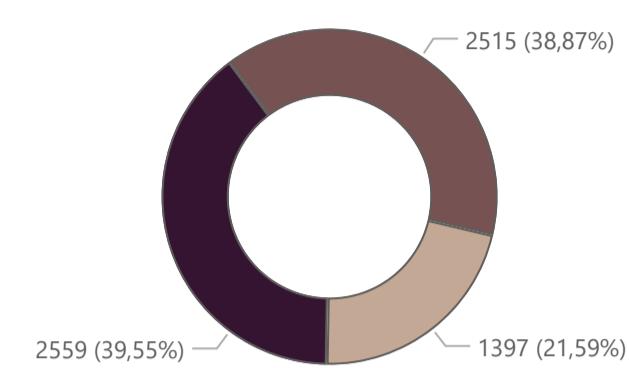
Avg Number of Orders per Account

2

Avg Amount of Orders per Account

5 649

Number of Orders - per Users (includes Owners and Disponents)



ORDERS INFO

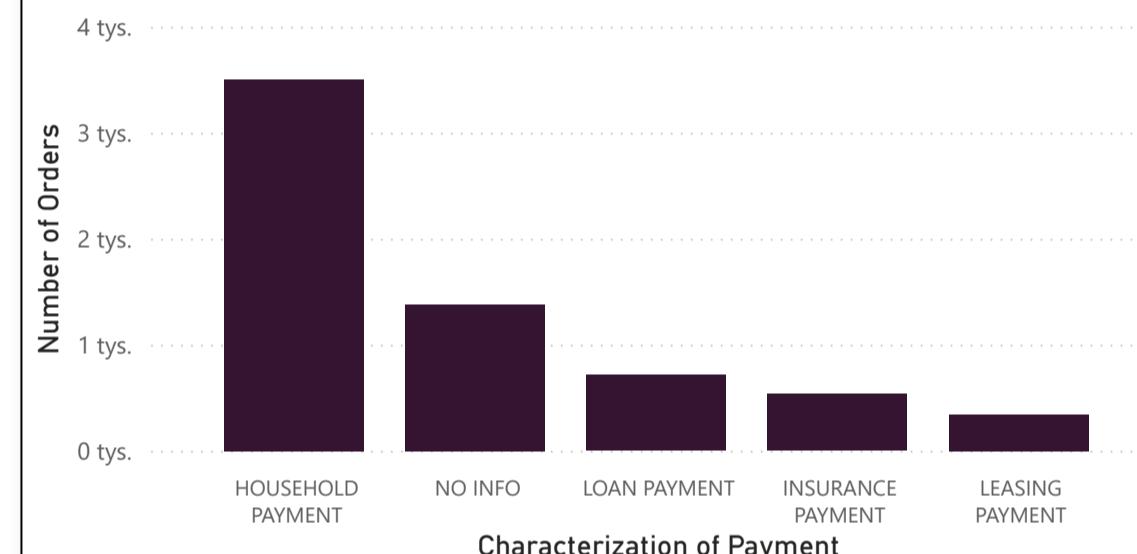
Total Number and Amount of Orders

Characterization of the Payment

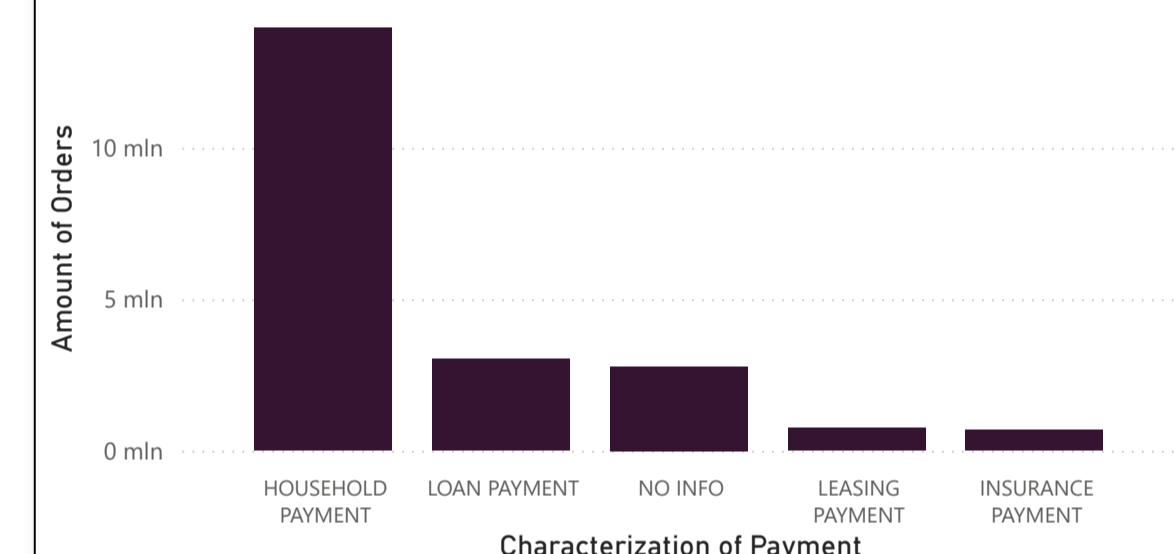
Number of Orders Total Amount of Orders

HOUSEHOLD PAYMENT	3502	13 965 417
NO INFO	1379	2 781 938
LOAN PAYMENT	717	3 035 185
INSURANCE PAYMENT	532	686 927
LEASING PAYMENT	341	759 527
Suma	6471	21 228 994

Total Number of Orders per Payment Type



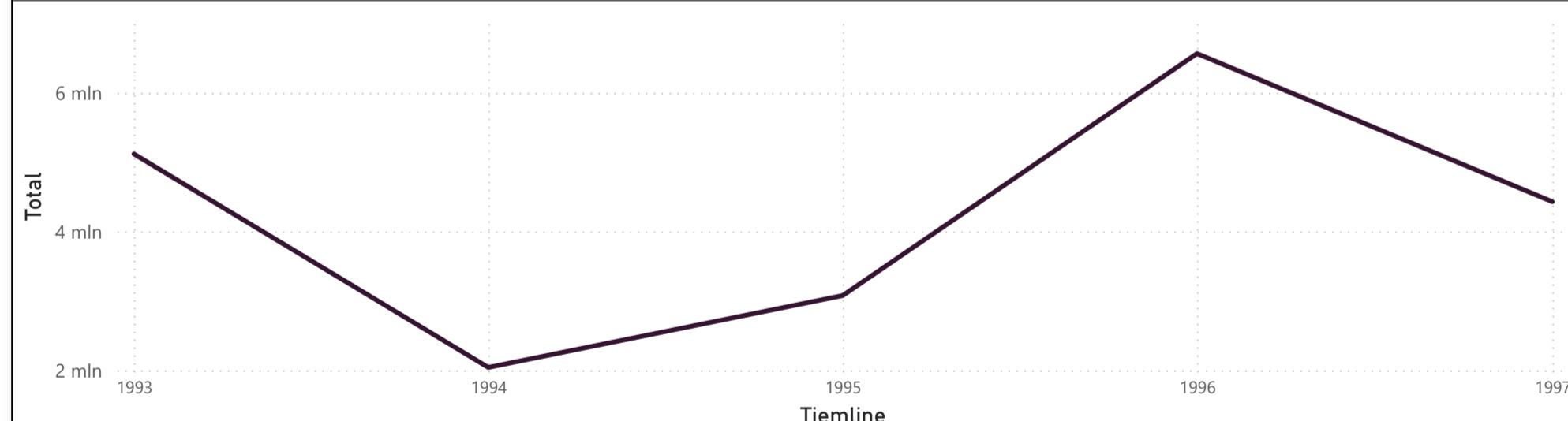
Total Amount of Orders per Payment Type



Amount of Orders

Number of Orders

Selected Metrics (Amount or Numer of Orders) in Time



Order Id

Wyszukaj

- 29401
- 29402
- 29403
- 29404
- 29405
- 29406
- 29407
- 29408

Account_id

Wyszukaj

- 1
- 10
- 100

ORDERS DETAILS

Orders details

Order Id	Account Id	Bank of the Recipient	Characterization of Payment	Order Amount
29401	1	YZ	HOUSEHOLD PAYMENT	2 452,00
29402	2	ST	LOAN PAYMENT	3 372,70
29403	2	QR	HOUSEHOLD PAYMENT	7 266,00
29404	3	WX	HOUSEHOLD PAYMENT	1 135,00
29405	3	CD	NO INFO	327,00
29406	3	AB	INSURANCE PAYMENT	3 539,00
29407	4	UV	HOUSEHOLD PAYMENT	2 078,00
29408	4	UV	HOUSEHOLD PAYMENT	1 285,00
29409	5	GH	HOUSEHOLD PAYMENT	2 668,00
29410	6	AB	HOUSEHOLD PAYMENT	3 954,00
29411	7	IJ	HOUSEHOLD PAYMENT	4 880,00
29412	8	UV	HOUSEHOLD PAYMENT	2 612,00
29413	8	IJ	NO INFO	6 712,00
29414	10	UV	HOUSEHOLD PAYMENT	7 033,00
29415	10	QR	LEASING PAYMENT	1 344,00
29416	11	ST	HOUSEHOLD PAYMENT	2 132,00
29417	12	CD	HOUSEHOLD PAYMENT	297,00