

## Data Manipulation and Cleaning work

### Account Table

1. Convert the Date attribute into a yyyy-mm-dd by adding 24 in year format in Excel or SQL  
=date(left(a1,2)+24,mid(a1,3,2),right(a1,2)).  
Date\_add("940612", Interval 24 Year).
2. Replace in frequency attribute "POPLATEK MESICNE" AS Monthly issuance, "POPLATEKTYDNE" AS weekly issuance, and "POPLATEK POBRATU" AS issuance after a transaction in Excel or create a case statement in SQL.

### CARD Table

1. Replace type attribute value "junior" as Sliver, "Classic" as Gold, And "Gold" as Diamond by using replace in Excel or by using update in SQL.
2. Convert issued attribute into yyyy-mm-dd adding 23 in year.

### CLIENT Table

1. Convert bith\_number attribute to yyyy-mm-dd format and also create another column named sex by applying in bith\_number 0 for females and 1 for males.  
(=if(mod(birth\_number,2)=0, "Female", "Male")) in excel or using case statement in SQL.  
=date(left(a1,2)+24,mid(a1,3,2),right(a1,2)).

### DISTRICT Table

1. Change all column names and **delete the attributes a12 and a13** and rename  
**a15 -> No\_committed\_crime\_2017**  
and **a16 -> No\_committed\_crime\_20178**

## 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 Table**

1. Convert the Date Attribute into **yyyy-mm-dd** format adding **23** in year.
2. Convert **Status Attribute** value

“A” -> Contract Finished,

“B” -> Loan Not Paid,

“C” -> Running Contract,

“D” -> Client in debt.