

## Data Manipulation and Cleaning work

### 1) Account Table

1. Convert the Date attribute into a yyyy-mm-dd by adding **24** in year format in Excel or SQL
  - 1993 -> 2017
  - 1994 -> 2018
  - 1995 -> 2019
  - 1996 -> 2020
  - 1997 -> 2021

In excel by using formula:

**=DATE (24+LEFT(D2,2), MID(D2,3,2), RIGHT(D2,2))**

(now our 5 years like 2020,21,22,23 and 24)

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.
3. Create a Custom Column Card\_Assigned and assign below:
  - Monthly issuance → Silver
  - Weekly issuance → Diamond
  - Issuance → Gold

### 2) CARD Table

1. Replace type attribute value “junior” as Silver, “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 26 in year.  
**=DATE (23+LEFT(D2,2), MID(D2,3,2), RIGHT(D2,2))**

### 3) CLIENT Table

1. Convert birth\_number attribute to yyyy-mm-dd format.
2. create another column named Gender by applying in birth\_number 0 for females and 1 for males.  
**(=if(mod(birth\_number,2)=0, “Female”, “Male”))** in excel or  
using case statement in SQL.

For **Male** its in **YYMMDD** format and for female it is **YYMM+50DD** for **Women**

#### 4) DISTRICT Table

1. Change all column names and delete the attributes **a12** and

##### 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

2. **a1→ district code**
- 3.

#### 5) LOAN Table

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

“A” as Contract Finished, “B” as Loan Not Paid, “C” as Running Contract, and “D” Client in debt.

## 6) Transactions Table

Do the following, where count is the highest sort it in descending order and change the year from 2022,2021,2020 and so on

--DATA TRANSFORMATION

/\*

2021 -> 2017

2020 -> 2018

2019 -> 2019 -- NO CHANGE

2018 -> 2020

2017 -> 2021

2016 -> 2022

UPDATE TRANSACTIONS

- SET BANK = 'Sky Bank' WHERE BANK IS NULL AND YEAR(DATE) = 2022;

UPDATE TRANSACTIONS

SET BANK = 'DBS Bank' WHERE BANK IS NULL AND YEAR(DATE) = 2021;

UPDATE TRANSACTIONS

SET BANK = 'Northern Bank' WHERE BANK IS NULL AND YEAR(DATE) = 2019;

UPDATE TRANSACTIONS

```
SET BANK = 'Southern Bank' WHERE BANK IS NULL AND YEAR(DATE)
= 2018;
```