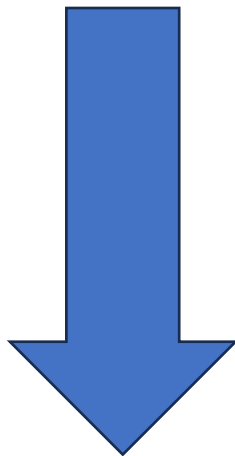


STEPS INVOLVED

- DATA CLEANING – downloaded raw csv files and cleaned the data in MS Excel including changing data types, switching values for specific features using functions and formulas like IF, SWITCH, DATE & other functions.
- DATA INGESTION – Ingested the clean csv files into Snowflake for table creation, assigning Primary & foreign keys, making of ER diagram.
- DATA TRANSFORMATION/MANIPULATION – Transforming the clean data by adding a few columns and performing joins operations between tables to generate KPI's for further analysis.
- CREATING CONNECTION - Connecting all the cleaned and transformed tables from Snowflake to Power BI and connecting the local server using ODBC connector.
- DATA VISUALIZATION – Visualizing the data using custom charts to draw insights which will help the organization in decision making.

STEPS FOR DATA CLEANING AND MANIPUATION



Data Manipulation and Cleaning work

Account Table

1. Convert the Date attribute into a yyyy-mm-dd by adding **23** in year format in Excel or SQL

- 1993 -> 2016
- 1994 -> 2017
- 1995 -> 2018
- 1996 -> 2019
- 1997 -> 2020

Converted the dates using text-to-column then using edate to add 23*12 months to the dates, (Alternate way – use date func with left/right function).

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.

Completed this task by using SWITCH func.

3. Create a Custom Column Card_Assigned and assign below :

- Silver -> Monthly issuance
- Diamond - weekly issuance
- Gold - Issuance after a transaction

Used if function to get the modification done

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.

Completed this task by using SWITCH func.

2. Convert issued attribute into yyyy-mm-dd adding 24 in year.

Using =DATE(YEAR(D3),MONTH(D3)-1,DAY(EOMONTH(D3,-1)))

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.

Converting the birth_number to birth_date using date, left, mid, right.

DISTRICT Table

1. Change all column names and delete the attributes **a12** and **a13**.
- 2.

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 24 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.

Used switch function to replace values.

use date func with left, right & mid function.

In th Transactions Table do the following , whosoever count is highest sort it in descending order and change the year from 2022,2021,2020 and so on

--DATA TRANSFORMATION

/*

2021 -> 2022

2020 -> 2021

2019 -> 2020

2018 -> 2019

2017 -> 2018

2016 -> 2017

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(TRANSACTION_DATE) = 2019;

UPDATE TRANSACTIONS

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

UPDATE TRANSACTIONS

SET BANK = 'ADB Bank' WHERE BANK IS NULL AND YEAR(TRANSACTION_DATE) = 2017;