# **Homework – Data Analyst (Data and Tools)**

The homework consists of programming tasks and data analysis tasks. The programming should be done in SQL language. You can use SQLite (<https://sqliteonline.com/>) for the SQL part. In addition, provide the solutions to the questions/analysis/problems with program of own choosing (e.g Word, Excel, Notebook). The task is based on two datasets, *defaults.csv* and *exchangerates.csv*. Each default is uniquely defined by AgreementGenID and DefaultDate.

*defaults.csv* contains variables:

AgreementGenId – Agreement identification number

DefaultDate – Date when default started

DefaultEndDate – Date when default status ended

DefaultRankNum – Ranking number for the default. The lower the riskier.

DefaultType – Default type code

ValutaKod – Currency code

LossAmount – Loss amount

EAD – Exposure at default

*exchangerates.csv* contains variables:

Currency – Currency name

CurrencyCd – Currency Code

ReportDate – Evaluation date

Valuation – Currency rate to SEK

The first task is to import these datasets and start using them for different subtasks brought below.

1. How many distinct defaults are in the dataset *defaults?*
2. Make a distribution graph of how many times agreements default.
3. Due to regulations, defaults that have less or equal to 9 months between the end date and new defaults start date must be merged into one with the first defaults start date and the latest end date. Perform this 9 month aggregation on default data with the following rules for other variables:
   1. The EAD and loss amount come from the first default.
   2. The DefaultTypeCd must be the riskiest one. The riskiest default is the one with the lowest ranking number.
4. What is the new number of unique defaults?
5. Analyse EAD-s and loss amounts.
6. Make a new binary variable that shows if the default is open now. How many defaults are closed and how many are open?
7. Add 1.6% of EAD to each loss amount. Find the sum of new loss amount and old. Also calculate the loss percentages on total level and on default type level with the new loss amount.
8. Analyse the data and describe the relevant/important characteristics.

Save the programming task as *.sql* file and the analysis file as you own choosing. Save your final data also as a *.csv* file. Save the complect as *firstname\_lastname.zip* file and send it back to us.