Please note that we do not expect the challenge to take more than 1–2 hours depending on how familiar you are with the tasks involved and different tools.

Deliverable:

1. You should submit one sql script– letting us see how and why you went about our challenge in a given way is integral to completing this task. The code should be efficient, clean and sufficiently commented.

2. The file with the result including tables and graphs. To make the task easier we recommend presenting results via Excel (but you can do it in any visual format). Please note that the task number 5 should be prepared in PowerBI and send us as a project.

Your submission should be in a .zip file – do not host it on GitHub or other platforms. Name the submission < your\_first\_name>\_<your\_last\_name>\_ht.zip. For example, jan\_kowalski\_ht.zip. Please do not protect it with a password.

We recommend to use <https://sqliteonline.com/> to upload files and use SQL queries for analysis. You will not need to install anything with “sqliteonline” web page, but you can also create your own local database.

**All deliverables should be in English. You need to use PostgreSQL.**

One of the most simple and popular consumer finance products is an Unsecured Personal Loan. Competition in the market is strong and margins are slim so the product has to be well prepared. A few months ago The Lender starts issuing loans for a totally new group of the customer. It’s the right time to analyze the result. Your task is to prepare high-level information about customer performance. You will find two tables with random sample data (due to reducing the size, but remember that in real life you are going to work with a huge amount of data).

Relevant files in data directory:

1. application.csv – table with applications information

* Application\_id
* Client\_id
* Application\_date\_created
* Decision – the final decision of approving or rejecting the application
* Limit – the final limit (amount of the loan) proposed to the customer

1. portfolio.csv – typical table for banking companies where you can see performance of loans (please note that in this file you can find only loans related to applications from )

* Application\_id
* Loan\_id
* Report\_date – Report data that shows actual situation of the loan for that particular day
* Client\_id
* loan\_date\_created
* term – term of the loan in days
* amount\_issued – amount issued to customer
* paid\_amount – amount paid by client (please note that the customer has to pay back the loan X days after date Y, where X is term and Y is loan data created)

**Tasks:**

1. Explore the data to get an understanding of the problem and how to proceed in the later steps.

2. Sale’s Manager would like to see:

* Testing period
* Number of applications
* Number of clients
* Approval Rate – relationship between approved and all applications
* Average limit
* Number of loans
* Issued rate – relationship between number of the loan and all applications
* Average issued amount

3. Portfolio Manager would like to add a new column to the table portfolio – DPD (Days Past Due; if you are not familiar with definition of DPD, please discover it on your own).

4. Risk Manager would like to have a new table for analysis on him/her own, where each row represents each application with next column:

* application\_id
* application date
* decision
* limit
* Is\_issued – flag that shows if the loan was created for that particular application
* Loan\_date
* Amount issued
* Maximum DPD
* DPD\_30 – flag that means that the DPD exceed 30 days
* default flag – where default means that the loan wasn’t paid back

5. CEO would like to see with visualization in PowerBI (summary table and graph in time):

* Number of applications
* Approval Rate
* Number of loans
* Issued rate
* DPD\_30 rate - relationship between flag DPD\_30 and number of loans
* Default rate - relationship between default flag and number of loans

**Good luck! 😊**