

## TASK 1

You have received a parquet file with leads delivered to one of our clients exported from the backend database. This client reports back to us on the progress of each lead through an SFTP server using CSV snapshots that he generates from his CRM system. Unfortunately, his reports are created manually and uploaded to the server by one of his employees. Moreover, the only links between both data sources are hashed emails and phone numbers of leads that we generated for them.

You have been asked to create a process that will match the data from our backend with the reports from our customers to help us keep improving the quality of generated leads. The most important aspect of lead quality is the conversion rate from the delivered leads to appointments with homeowners. Our client provides this information in column `'set'` that indicates the successfully set appointment and as a scheduled appointment data `'Appt Date'`. Furthermore, the client shares with us if the appointment ended with a demo, which is a next step of their usual sales process.

Create a validated dataset that can be used for further lead quality analysis. Use any Python and/or SQL-based tools that will allow you to achieve desired results. Share with us the output data and all relevant code.

## TASK 2

Create a DBT project that will define the source table that you produced in Task 1 and build a model that will create a summary table showing how many leads we have generated per month within each state, what percentage of them has converted to appointments, and then to demos. Ensure that your project is production ready.

As a solution, please provide the link to a git repository containing DBT project files.