Data analyst case study

There are multiple CSV files essentially acting as simple tables from a database.

File “trans” contains data about all deposits that happened during the given period with the following columns:

- login – account login number
- created_at – datetime of the transaction. Timezone can be ignored for the purposes of this case study
- transaction_id – ID of the transaction
- amount – amount of the transaction in account currency
- currency – account currency

File “trader” contains data about individual trading accounts with the following columns:

- login – account login number where the relationship between “login” and “transaction_id” is 1:N
- client – client identifier
- account_id – another ID describing the account from an external system
- first_deposit_id – transaction_id of the first deposit on that specific login

File “clients” contains data about clients with the following columns:

- client – client identifier where the relationship between “client” and “login” is 1:N
- client_id – another ID describing the client from an external system
- _created_on – datetime of when the client profile was created
- type – registration type. “Full” and “Light” are live registrations, “Demo” is demo

File “partner codes” contains a list of parter codes along with a unique partner_code_id assigned to them which is just an auto-increment value. The actual partner code is in the column “code”.

File “a2p” is a reference table between accounts and partner codes.

- account_id is ID from “trader”
- ref_id is just the reference ID
- partner_code_id is reference from “partner codes”

The goal of this study is to provide information about first time deposits (FTD) over time on client level. FTD is calculated as the first deposit the client made across all of their logins.

For each client the date of their FTD should be calculated and from these dates a chart should be created. There should be two charts – one showing just the number of FTDs per each month and the other showing the amount of USD deposited in the first deposit per each month. The data should be filtered by the clients.type column as only “Full” and “Light” types are relevant.

As the last output (either on the chart of as separate output), we would like to know how many FTDs we can expect in the next month.

The whole solution should be created in Python.