DATA NEXUS HUB WEEKLY COMPETITION

Background:

In the realm of banking, personal loans serve as a significant revenue source, with a typical interest rate of 10% for a two-year loan in the UK. In September 2022 alone, UK consumers borrowed £1.5 billion, generating an estimated £300 million in interest for banks over two years.

Tasks:

Mogul Bank, a UK-based bank has tasked a data professional with cleaning and storing the data they collected as part of a recent marketing campaign, which aimed to get customers to take out a personal loan. Mogul stakeholders are also interested in knowing key helpful insights about their marketing data.

As the Bank aim to conduct more campaigns in the future, they'd love you to implement a robust RDBMS database capable of accommodating data from future campaigns, with a schema comprising three tables: Client, Campaign, and Economics.

Lastly, you will write the SQL code that the bank can execute to create the tables and populate with the data from the csv files. As the bank are quite strict about their security, you'll save SQL files as multiline string variables that they can then use to create the database on their end.

Data and Tools:

They have supplied you with a csv file called "bank_marketing.csv", which you will need to clean, reformat, and split, to save separate files based on the tables you will create. It is recommended to use pandas and Jupyter notebook, or Python for these tasks.

Data Dictionary:

client

column	data type	description	original column in dataset
id	serial	Client ID - primary key	client_id
age	integer	Client's age in years	age
job	text	Client's type of job	job
marital	text	Client's marital status	marital
education	text	Client's level of education	education
credit_default	boolean	Whether the client's credit is in default	credit_default
housing	boolean	Whether the client has an existing housing loan (mortgage)	housing
loan	boolean	Whether the client has an existing personal loan	loan

campaign

column	data type	description	original column in dataset
`campaign_id`	`serial`	Campaign ID - primary key	N/A - new column
`client_id`	`serial`	Client ID - references `id` in the `client` table	`client_id`
`number_contacts`	`integer`	Number of contact attempts to the client in the current campaign	`campaign`
`contact_duration`	`integer`	Last contact duration in seconds	`duration`
`pdays`	`integer`	Number of days since contact in previous campaign (`999` = not previously contacted)	`pdays`
`previous_campaign_contacts`	`integer`	Number of contact attempts to the client in the previous campaign	`previous`
`previous_outcome`	`boolean`	Outcome of the previous campaign	`poutcome`
`campaign_outcome`	`boolean`	Outcome of the current campaign	`y`
`last_contact_date`	`date`	Last date the client was contacted	A combination of `day` , `month` , and the newly created `year`

economics

column	data type	description	original column in dataset
`client_id`	`serial`	Client ID - references `id` in the `client` table	`client_id`
`emp_var_rate`	`float`	Employment variation rate (quarterly indicator)	`emp_var_rate`
`cons_price_idx`	`float`	Consumer price index (monthly indicator)	`cons_price_idx`
`euribor_three_months`	`float`	Euro Interbank Offered Rate (euribor) three month rate (daily indicator)	`euribor3m`
`number_employed`	`float`	Number of employees (quarterly indicator)	`nr_employed`

FIND THE DATASET ON THE MICROSOFT TEAMS OR WITH YOUR GROUP LEAD