



Personal loans are a lucrative revenue stream for banks. The typical interest rate of a two-year loan in the United Kingdom is [around 10%](#). This might not sound like a lot, but in September 2022 alone UK consumers borrowed [around £1.5 billion](#), which would mean approximately £300 million in interest generated by banks over two years!

You have been asked to work with a bank to clean the data they collected as part of a recent marketing campaign, which aimed to get customers to take out a personal loan. They plan to conduct more marketing campaigns going forward so would like you to ensure it conforms to the specific structure and data types that they specify so that they can then use the cleaned data you provide to set up a PostgreSQL database, which will store this campaign's data and allow data from future campaigns to be easily imported.

They have supplied you with a csv file called `"bank_marketing.csv"`, which you will need to clean, reformat, and split the data, saving three final csv files. Specifically, the three files should have the names and contents as outlined below:

`client.csv`

column	data type	description	cleaning requirements
<code>client_id</code>	<code>integer</code>	Client ID	N/A
<code>age</code>	<code>integer</code>	Client's age in years	N/A
<code>job</code>	<code>object</code>	Client's type of job	Change <code>"."</code> to <code>"_"</code>
<code>marital</code>	<code>object</code>	Client's marital status	N/A
<code>education</code>	<code>object</code>	Client's level of education	Change <code>"."</code> to <code>"_"</code> and <code>"unknown"</code> to <code>np.NaN</code>
<code>credit_default</code>	<code>bool</code>	Whether the client's credit is in default	Convert to boolean data type
<code>mortgage</code>	<code>bool</code>	Whether the client has an existing mortgage (housing loan)	Convert to boolean data type

campaign.csv

column	data type	description	cleaning requirements
client_id	integer	Client ID	N/A
number_contacts	integer	Number of contact attempts to the client in the current campaign	N/A
contact_duration	integer	Last contact duration in seconds	N/A
previous_campaign_contacts	integer	Number of contact attempts to the client in the previous campaign	N/A
previous_outcome	bool	Outcome of the previous campaign	Convert to boolean data type
campaign_outcome	bool	Outcome of the current campaign	Convert to boolean data type
last_contact_date	datetime	Last date the client was contacted	Create from a combination of <code>day</code> , <code>month</code> , and a newly created <code>year</code> column (which should have a value of <code>2022</code>); Format = "YYYY-MM-DD"

economics.csv

column	data type	description	cleaning requirements
client_id	integer	Client ID	N/A
cons_price_idx	float	Consumer price index (monthly indicator)	N/A
euribor_three_months	float	Euro Interbank Offered Rate (euribor) three-month rate (daily indicator)	N/A