# Financial Crime Data Scientist position technical assessment

## SQL Proficiency

We’re providing you with synthetic daily exchange rate information between GBP and various foreign currencies, and a sample of transacting days. We have some days for which we’re missing the exchange rates for some transactions.

• How would you fill the missing values?  
• Can you propose a solution that uses the last existing exchange rate for that currency pair?  
• Propose a solution using SQL-lite.

The solution should only use SQL constructs and not rely on libraries from other languages.

## Data Science case study

In this case study you are provided with a simple synthetic but realistic dataset of trades in csv format covering data from 2018/12 to 2020/01 and asked to perform analysis and a modelling task on it to predict if a trade is suspicious. A transaction could be classified as suspicious for different money laundering suspicion reasons. The description of the data variables is as below:

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| Source System | The source system from the data |
| Leg Type | The type of the leg in the trading strategy |
| Trade Date | The date of the trade in dd/mm/yyyy format |
| Product Name | The name of the product |
| Product Group | The group of the product |
| Notional Amount Currency | The currency of the underlying asset |
| CountryCodeOffice | The country code of the office |
| ML Risk Rating | The rating of the Money Laundering Risk |
| Client Name | The name of the client |
| Notional Amount | The amount of the underlying asset |
| Gbp Notional Amount | The amount of the underlying asset in £ |
| Fx\_rate | The rate of £ amount against the notional amount |
| Flag | Is Suspicious |

Based on this dataset, you are required to develop a Machine learning solution to address the questions below using Python (this shouldn’t take you more than a few hours):

1. Perform a basic exploratory analysis and describe the quality of the data
2. Engineer a small number of features
3. Develop a machine learning solution to help identify future suspicious trades
4. Provide an evaluation of your model, explain what metrics you have used, what parameters are used by your model, and any limitations you can think of.
5. Do best attempts at building a clean data pipeline providing the basis for production code
6. Demonstrate using logging to help debugging in the future
7. Build at least one unit-test and one validation-test.

We’re not expecting you to write production ready solution in a few hours, but we care about your adopting programming best practices as much as about a sound data science methodology. Make clear which areas of concern in your solution would need to be improved.

Please send your code/notebook with any visualisation and results you have produced within 7 days after you received this email. Please message [michael.heusch@santander.co.uk](mailto:michael.heusch@santander.co.uk) or [max.palmer@santander.co.uk](mailto:max.palmer@santander.co.uk) if you need clarification on any of the above and to submit your final solution.