



## Tech Challenge - Fraud Data Analyst

---

As SwissBorg grows and establishes itself in the market, it becomes increasingly appealing to some users that wish to abuse the neat experience offered by the app. These users exchange funds often obtained through illicit activities, and use the platform as a gateway to convert fiat to crypto or vice versa. Their main objective is to make it harder for authorities to trace back the source of the funds.

In this tech challenge, we provide you two datasets, containing user and transaction data. Your goal is to perform an exploratory analysis to investigate potential patterns in the fraud cases, and how they could be distinguished from legitimate users, using the questions below as a guide. Please prepare a short report or notebook summarizing the findings, and attach any code you have used to come up with your conclusions.

Remember that, at the same time as we want to identify all suspicious cases, we need to be careful to limit the amount of false positives, since these can have negative consequences for the user experience and the support team's workload.

- 1) Create a set of up to 6 charts that best describe the evolution of fraud cases over time in the dataset. Imagine that these charts would form a dashboard.
- 2) What are the 3 main differences between fraud and legitimate users?
- 3) Is there a distinctive transaction behaviour of fraud users?
- 4) Write one SQL query that retrieves user\_ids with that followed this pattern (in the sequence described below):
  - a) Make one or more fiat deposits summing at least 2k EUR. The first fiat deposit is within 4 days after the onboarding. These are the first deposits of the user, with no crypto deposits before.
  - b) Exchange most of the funds ( $\geq 90\%$ ) to ETH or BTC in one transaction
  - c) Withdraw at least 2k EUR of ETH or BTC within 4 days after the first fiat deposit. This crypto withdrawal is the first withdrawal by the user.

TIP: You can use the python library pandasql to work on the query using pandas dataframes as source.

**Good luck and be awesome!**

## Data dictionary

user\_info.csv

Column	Description
user_id	The user's unique identifier
age	The age of the user in years
country_code	The user's country
was_referred	Boolean indicating if the user was referred in the app through the member-get-member program
onboarding_completed_at	Timestamp when the user completed the onboarding
t2_upgrade_at	Timestamp when the user upgraded to Tier 2. This entails a photo ID + address verification, and allows the user to deposit up to 50K EUR and withdraw 1M EUR
t3_upgrade_at	Timestamp when the user upgraded to Tier 3. This entails a origin of funds verification, and removes deposit or withdrawal limits
became_premium_at	Timestamp when the user became premium
premium_tier	Current premium tier of the user
is_fraud	Boolean indicating if the user has been flagged as suspicious by the compliance team

transaction\_info.csv

Column	Description
user_id	The user's unique identifier
transaction_id	The transaction's unique identifier
timestamp	The transaction's timestamp
event_kind	<p>The type of transaction. Possible values: fiat_deposit / fiat_withdrawal / crypto_deposit / crypto_withdrawal / exchange</p> <p>Fiat means "traditional" currencies such as EUR, USD, GBP etc, whereas Crypto means tokens such as BTC, CHSB, ETH etc.</p> <p>Deposit means transferring funds from an external account to SwissBorg, and withdrawal means the opposite. Exchange means trading one currency to another within the app.</p>
currency	The currency of the transaction. In case of exchanges, it contains the format FromCurrency_ToCurrency. For instance, if BTC_USD, it means that the user traded BTC to USD
amount_in_eur	The equivalent EUR amount of the transaction