Task 2: Catching Fraud

2.1)

The query displays how much money a user has spent in total with merchants, grouped by country,
where users have paid via GAIA
and
where the user's first registered phone number by country matches the merchant's country. In other words, the payments a user made in their home country.

The results are displayed in €* in a descending order.

* Converted from payed currency into €.

2.2)

```
SELECT USER_ID
FROM (SELECT Min(CREATED_DATE), USER_ID, TYPE, STATE, CURRENCY, AMOUNT
FROM transactions
GROUP by USER_ID), fx_rates
WHERE TYPE = 'CARD_PAYMENT' AND STATE = 'COMPLETED' AND CURRENCY = ccy AND base_ccy = 'USD' AND (AMOUNT * rate) >= 10;
```

I have stored the output in an according file. See solution.txt. Also the corresponding sqlite database I tested the query on can be found as a file named test.db.

2.3)

I had several approaches to this task, but none of them proved to be practical.

1. I could have trained a predictive model. Logistic regression or XGBoost would have come into question. However, there were not enough data sets available for both.

2. I could have created personas. The characteristics of the persona would be based on average/mean values of the respective attributes. E.g. age, KYC, etc.…

Also time restrictions limited my possibilities to come up with a creative idea.