

## Data Scientist – Recruitment Exercise

### Background

A fictional website 'sell-your-stuff.com' allows clients to buy and sell items that are not being used anymore (chairs, laptops, books, etc.). Some of the clients make their living by buying products refurbishing and then reselling it. But most are only interested in selling things they don't use anymore or buying cheap stuff.

The process is quite simple, a client creates an account (choosing one of the possible account types, each account type follows specific rules and legislations for different types of products and countries). One client can have only one account.

After account creation clients need to deposit money to buy/sell products, as the website itself works as a broker and guarantees that both clients will be satisfied.

Internally the company keeps information of every product bought and sold splitting them into categories and socioeconomic factors. This data is then sold to third parties following data protection legislations (no personal information is disclosed). As a rule of thumb, the company sells the data by roughly 1% of the total clients have bought. (e.g. a client that buys US\$ 100.00 worth of products will generate US\$ 1.00 to the company ).

As the information contained on these tables can reach terabytes, summary tables were created in order to help data analysts and managers to take decisions regarding marketing strategies and resources allocation. The most used tables are:

- *client*: Basic anonymized information about clients and client's account.
- *transactions*: Summary of transactions amounts of each account aggregated daily.
- *campaigns*: High level description of past marketing campaigns, dates and budget.

- A complete description of the columns can be found on Appendix 1. - The SQLite3 database (*company.db*) attached contains a sample of the data for the

summary tables and should be used to answer the following question:

## Question

The management of sell-your-stuff.com wants to set up marketing strategies for the next fiscal year based on the summary tables. Your job is to analyse the performance of past campaigns, coupons, profit by countries, etc. and make a report showing information to support your insights.

## Appendix

### 1. Database specification

- **Clients:**

- *account*: Unique index linked to an account (1 client -> N accounts).
- *type*: Type of the account.
- *residence*: Country of residence of the client (ISO 3166-1 alpha-2).
- *date\_joined*: Date of account creation.
- *indication\_coupon*: If the client was indicated by some of the partners this value reflects the partner's ID.
- *first\_deposit\_amount*: Amount deposited on client's first deposit (in US dollar).
- *first\_deposit\_date*: Date of first deposit.
- *first\_transaction\_amount*: Amount of client's first transaction (in US dollar).
- *first\_transaction\_date*: Date of first transaction.
- *balance\_amount*: Current balance.

- **Transactions:**

- *transaction\_date*: Date (ISO 8601 up to day).
- *account*: Index to connect to client's account.

- *total\_buy*: Total spend on the day buying products (in US dollar).
- *total\_sell*: Total earned on the day selling products (in US dollar).
- *total\_deposits*: Total deposited on the day (in US dollar).
- *total\_withdrawals*: Total withdrawn on the day (in US dollar).
- *count\_contracts*: Number of items sold on the day (each item represents a different contract).
- *count\_deposits*: Number of deposits on the day (a client can deposit multiple times on a day).
- *count\_withdrawals*: Number of withdrawals on the day (a client can withdraw multiple times on a day).

- **Campaigns:**

- *start\_date*: Date in which the company started the campaign (ISO 8601 up to day).
- *end\_date*: Date in which the company finished the campaign (ISO 8601 up to day).
- *total\_spend*: Total spend for the campaign (including all expenses, in US dollar)
- *country*: Country in which the campaign was focuses (ISO 3166-1 alpha-2), Null values represents global campaigns.