AgTech Use Case

Disclaimer: this case is based on a made-up dataset that mimics challenges faced by data analysts at Yara.

Challenge: the candidate is assigned the tasks below and shall return:
- Code sheet (i.e., SQL, python, python Notebook, R, etc.);
- This file with answers to tasks 1 and 2;
- A single presentation deck for tasks 3 and 4, focused on a business audience.

The Digital Innovation organization at Yara has developed a digital application that goes to farmers' hands and support them in the process of taking decisions on how to best apply fertilizers and when. They contribute to an efficient use of fertilizers and therefore can impact on crop yield, return on investment, while also driving a nature-positive outcome.

The success of Yara's app is measured against the following KPIs:
- Active users: users that have used the app at least once in the last 12-rolling months. If last event is earlier than that, the user is considered to have churned.
- Active fields: non-archived fields belonging to active users.
- Active hectares: field area belonging to active fields.

Recently, a loyalty program was created, and, among others, it encourages farmers to use the app extensively to draw the boundaries of their fields, explore the created fields (use the app to view created fields and configure them), monitor them (seeing climate, satellite images, etc.), create fertilizer application maps, and crop nutrition recommendations. To become a member, the farmers receive vouchers from Yara's commercial units, which also allows for measuring the engagement of those units.

Tasks

1) Before beginning any type of analysis, usually it is necessary to create a SQL query to build a dataset. Using the tables from the schema file, create a query that returns the following columns:
   - User id (yara_user_id);
   - Business unit (bu_name);
   - Employee (employee_name);
   - Promo-code (promo_code);
   - Date of registration (user_reg_date);
   - Total number of fields;
   - Total number of active fields;
   - Total field area;
   - Total active field area.

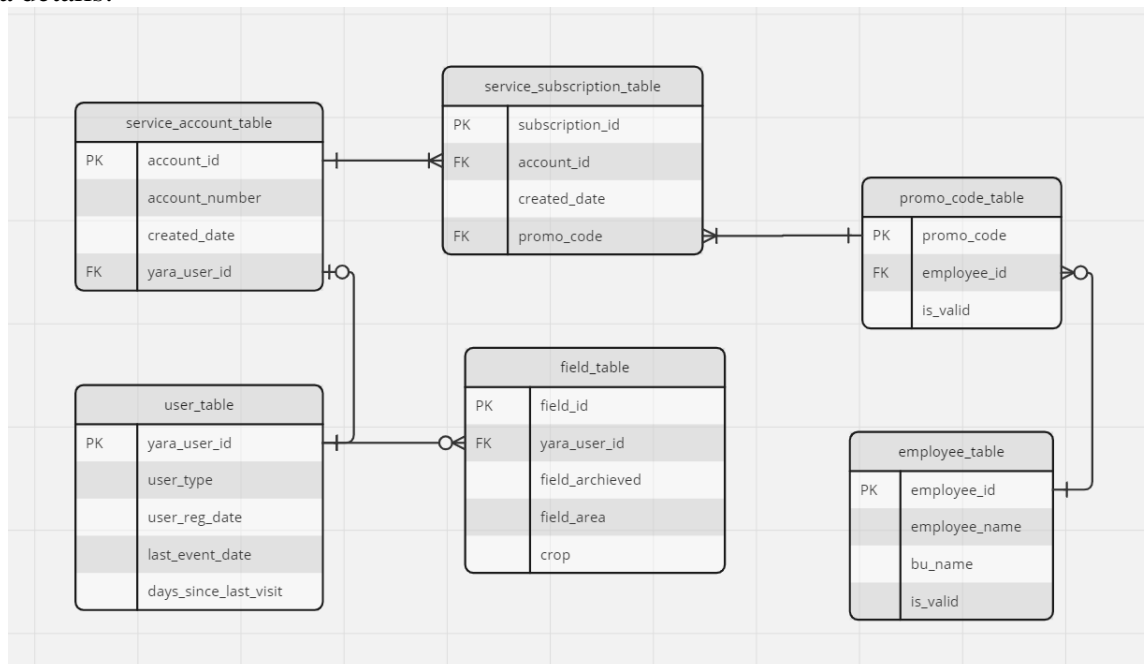Important: The resulting dataset should have only one row per user.

Schema details:

Figure 1: Schema details

user_table: Each row represents a user, with a unique identifier (*yara_user_id).*
field_table: Each row represents a field, each field is related to one user, but a user can have multiple fields.
promo_code_table: Each row represents a promo-code (each promo-code is unique) that loyalty program users need to register in the app.
employee_table: Each row represents an employee (salesperson). It contains information about the BU the employee is allocated.
service_account_table: This table is from a service that Yara uses to generate promo-codes to promote the usage of our tools, here the details from the account of this service are located. Each row represents an account that is used to generate the promo-codes. Unique account identifier is the *account_id.*
service_subscription_table: The same service provides a subscription of each account to certain programs, in this table the promo-codes to each of such initiatives are stored. Each row represents a subscription from a user. The *account_id* field specifies the account associated with that subscription.

2) Using the dataset provided, answer the following questions:

A – How many days are there between the registration and last event from users in CC-BGX, CC-GUI and CC-PAM?

B – Please provide the values for the three previously explained KPIs, segmented by *user_type*. The reported values should include the metric in absolute value and it's percentual contribution (e.g., 100 ha / 32%).

3) The VP of Commercial Excellence Americas, the loyalty program's sponsor, requested the analytics team to evaluate the program's success by comparing loyalty to non-loyalty farmers. Are loyalty users different from regular users in relation to the main KPIs, app utilization and

churn? Another important aspect is to understand if there are any BUs being more successful in acquiring and maintaining users in our digital tools. Create a 10-15 min presentation, bearing in mind she is not familiarized with technical data analysis terms. Please feel free to suggest data driven action plans.

4) Build a user churn prediction model and propose an action plan to prevent them from becoming inactive. Create a 5-10 min presentation.

Answers:

1-

```sql
SELECT
    users.yara_user_id as user_id
    ,emp.bu_name as business_unit
    ,emp.employee_name as employee
    ,promo.promo_code as promo_code
    ,users.user_reg_date as date_of_registration
    ,count(distinct field.field_id) as total_number_of_fields
    ,sum(case
            when field.field_archived = 1 then 0
            else 1
        end) as total_number_of_active_fields
    ,sum(field.field_area) as total_field_area
    ,sum(case
            when field.field_archived = 1 then 0
            else field.field_area
        end) as total_active_field_area
FROM user_table as users
    INNER JOIN field_table as field ON field.yara_user_id = users.yara_user_id
INNER JOIN service_account_table as serv_acc ON serv_acc.yara_user_id = users.yara_user_id
    INNER JOIN service_subscription_table as sub ON sub.account_id = serv_acc.account_id
    INNER JOIN promo_code_table as promo ON promo.promo_code = sub.promo_code
    INNER JOIN employee_table as emp ON emp.employee_id = promo.employee_id
GROUP BY
    users.yara_user_id
```

2-
A) How many days are there between the registration and last event from users in CC-BGX, CC-GUI and CC-PAM?
The average number of days between registration and the last event for users in the CC-BGX, CC-GUI, and CC-PAM business units is 599.05

Please provide the values for the three previously explained KPIs, segmented by user_type. The reported values should include the metric in absolute value and it's percentual contribution (e.g., 100 ha / 32%).

The values for the three KPIs, segmented by user_type, are as follows:

user_type	Active Users	Active Users %	Active Fields	Active Fields %	Active Hectares	Active Hectares %
advisor	1,058	17.33%	41,364,734	96.92%	389,821,233.39	86.21%
farmer	5,047	82.67%	1,314,043	3.08%	62,332,950.03	13.79%

For the user_type "advisor," there are 1,058 active users, which account for 17.33% of the total active users. The number of active fields is 41,364,734, contributing to 96.92% of the total active fields. The active hectares for advisor users are 389,821,233.39, representing 86.21% of the total active hectares.

On the other hand, for the user_type "farmer," there are 5,047 active users, making up 82.67% of the total active users. The number of active fields is 1,314,043, accounting for 3.08% of the total active fields. The active hectares for farmer users are 62,332,950.03, representing 13.79% of the total active hectares.