Lead Conversions Tech Task

1. Context

A new report has been requested by product to help stakeholders analyse Lead Conversions in their gyms. The model will enable end users to gain insights into the effectiveness of different conversion strategies across different conversion types. Below describes some key business definitions:

> Lead: A potential member who has shown interest in joining the gym or utilising its services. When a user is created they automatically start as a LEAD until they purchased a membership or credit pack product.

> Client: If the lead is satisfied with their experience, they are then converted into a paying member aka CLIENT. This involves signing a membership agreement or purchasing a credit pack.

> Client Conversion Event: The event which triggers the conversion from lead to client. examples: MEMBERSHIP, USER_CREDIT.

1.1 Prerequisites and Project Set-up

You may approach this assignment using either SQL or Python. Below are the respective set-ups for both.

SQL Approach

A containerised version of PostgreSQL 16 is provided pre-loaded with all required tables and data. To successfully run the container, docker should be installed along with an appropriate database client to interact with the database.

To run and connect to the database run the following commands:

```
docker build -t glofox-lead-challenge-db ./
docker run -d --name my-postgresdb-container -p 5432:5432 glofox-lead-challenge-db
```

The running database can be connected to using the following credentials:

- **URL**: jdbc:postgresql://localhost:5432/glofox
- **USERNAME**: postgres
- **PASSWORD**: docker

Python Approach

The `data` folder in this project contains all necessary source data required.

Python 3.9 or higher should be installed.

Any additional packages used in your solution should be included, by you, in a `requirements.txt` file.

Workings should be added to the `main.py` file, and any additional files should be clearly documented and included in your submission.

Please include CSV outputs for each of your answers

1.2 Part 1

Using the provided input data create a query that will populate the `fct_client_conversion_events` table (See 1.6.1) with the details of first credit pack purchased and first membership purchased per user. Fields prefixed with `client_conversion_` should contain the details of either the first credit pack or membership depending on which was purchased first.

NOTE: Input data may be incomplete. Please take note of this and design your queries to handle these scenarios.

1.3 Part 2

After further feedback was gathered it turns out that some gyms only consider membership purchases as qualifying events for conversion while others only consider credit pack purchases as a conversion while the majority of the gyms still consider both events valid. Users want a column, `client_conversion_event_filter`, that they can use to filter the report so that it will show them the conversion details in either of the three possible scenarios.

To support this, update your query so that, for each user, up to 3 records are created:

1. One that only considers the earliest membership purchase with a column called `client_conversion_event_filter` with value `MEMBERSHIP`
2. One that only considers the earliest credit pack purchases with a column called `client_conversion_event_filter` with value `USER_CREDIT`
3. One that considers both the earliest memberships and earliest credit pack purchases with a column called `client_conversion_event_filter` with value `ALL`

The new model should adhere to the following rules:

- If the client purchased a membership AND a credit pack then create a `MEMBERSHIP` record, a `USER_CREDIT` record and an `ALL` record.
- If the client purchased a membership but DID NOT purchase a credit pack then create a `MEMBERSHIP` record and an `ALL` record.
- If the client DID NOT purchased a membership but DID purchase a credit pack then create a `USER_CREDIT` record and an `ALL` record.

Once you have the dataset of users and their first conversion event details construct a new model which will enable this client conversion filtering. Results should appropriately adhere to the expected schema, `fct_lead_conversions` (See 1.6.2)

NOTE: If you find that you have not been able to complete Part 1 (1.2) you can attempt Part 2 (1.3) using the table, `fct_client_conversion_events_part_2` (See 1.6.3) which has been pre-populated with data similar to what would be expected from Part 1.

1.4 Requirements

SQL Approach

Please complete the task using the PostgreSQL 16 container provided.

All SQL written should be clearly documented throughout.

Submit solution in an `.sql` file separate to the `setup.sql` that can be run against the container without requiring an modification.

Creation of intermediate tables is permitted if seen as necessary to complete the solution whether to improve readability or performance.

Please make an effort to complete each step to the best of your ability.

Python Approach

All Python written should be clearly documented throughout.

Submit solution including any additional libraries or packages used in a `requirements.txt` or any similar dependency management tool.

Please include CSV outputs for each part.

Please make an effort to complete each step to the best of your ability.

1.5 Input Schema & Definitions

dim_branch (figure 1.5.1)

Holds the branch related data.

Col Name	Type	Description
branch_id	VARCHAR	Identifier for the branch.
name	VARCHAR	Name of the branch.
email	VARCHAR	Email used by the branch.
city	VARCHAR	City where the branch is located.
country	VARCHAR	Country where the branch is located.
timezone_id	VARCHAR	Timezone identifier of the branch.
created_at	TIMESTAMP	Time when the branch was created at (UTC).

dim_user (figure 1.5.2)

Holds the user related data.

Col Name	Type	Description
user_id	VARCHAR	Identifier for the user.

Col Name	Type	Description
branch_id	VARCHAR	Identifier for the branch.
name	VARCHAR	Name of the user.
email	VARCHAR	Email used by the user.
created_at	TIMESTAMP	Time when the user was created at (UTC).

fct_credit_pack_purchases (figure 1.5.3)

Holds the data relating to credit pack products purchases.

Col Name	Type	Description
user_id	VARCHAR	Identifier for the user.
branch_id	VARCHAR	Identifier for the branch.
credit_pack_id	VARCHAR	Identifier for the purchased credit pack associated with the user.
credit_pack_purchased_at	TIMESTAMP	Time when the credit pack was purchased at (UTC).
created_at	TIMESTAMP	Time when the credit pack was created at (UTC).
credit_pack_purchase_details	JSON	Json object of credit pack purchased details.

credit_pack_purchase_details (figure 1.5.4)

Json object of credit pack purchased details.

JSON Key	Type	Description
name	VARCHAR	Name of the credit pack purchased.
source	VARCHAR	Source of the credit pack purchased.

fct_membership_purchases (figure 1.5.5)

Holds the membership purchased related data.

Col Name	Type	Description
user_id	VARCHAR	Identifier for the user.
branch_id	VARCHAR	Identifier for the branch.
user_membership_id	VARCHAR	Identifier for the purchased membership associated with the user.
credit_membership_purchase_atpack_purchased_at	TIMESTAMP	Time when the membership was purchased at (UTC).

Col Name	Type	Description
created_at	TIMESTAMP	Time when the membership was created at (UTC).
membership_purchase_details	JSON	Json object of membership purchased details.

membership_purchase_details (figure 1.5.6)

Json object of membership purchased details.

JSON Key	Type	Description
name	VARCHAR	Name of the membership purchased.
source	VARCHAR	Source of the membership purchased.

1.6 Out Schemas & Definitions

fct_client_conversion_events (figure 1.6.1)

Dataset capturing users and their client conversion event along with details associated with their first membership or credit pack event.

Col Name	Type	Description
user_id	VARCHAR	Identifier for the user.
branch_id	VARCHAR	Identifier for the branch.
local_user_created_at	TIMESTAMP	Time when the user was created at (Local Time).
lead_status	VARCHAR	Lead status of the user either LEAD or CLIENT
client_conversion_event_type	VARCHAR	Event type of the conversion either MEMBERSHIP or USER_CREDIT
client_conversion_event_id	VARCHAR	Identifier of the conversion event
client_conversion_event_local_created_at	TIMESTAMP	Time when the client conversion event was created at (Local Time).
client_conversion_event_name	VARCHAR	Name of the client conversion event
client_conversion_event_source	VARCHAR	Source of the client conversion event
first_user_membership_id	VARCHAR	Identifier of the purchased membership associated with the first event found for a given user.

Col Name	Type	Description
first_local_membership_purchased_at	TIMESTAMP	Time when the membership was purchased at (Local Time) associated with the first event found for a given user.
first_membership_name	VARCHAR	Membership name associated with the first event found for a given user.
first_membership_source	VARCHAR	Membership source associated with the first event found for a given user.
first_credit_pack_id	VARCHAR	Identifier of purchased credit pack associated with the first event found for a given user.
first_local_credit_pack_purchase_at	TIMESTAMP	Time when the credit pack was purchased at (Local Time) associated with the first event found for a given user.
first_credit_pack_name	VARCHAR	Credit pack name associated with the first event found for a given user.
first_credit_pack_source	VARCHAR	Credit Pack source associated with the first event found for a given user.

fct_lead_conversions (figure 1.6.2)

Dataset capturing users and their first client conversion event

Col Name	Type	Description
user_id	VARCHAR	Identifier for the user.
branch_id	VARCHAR	Identifier for the branch.
local_user_created_at	TIMESTAMP	Time when the user was created at (Local Time).
lead_status	VARCHAR	Lead status of the user either LEAD or CLIENT
client_conversion_event_type	VARCHAR	Event type of the conversion either MEMBERSHIP or USER_CREDIT
client_conversion_event_id	VARCHAR	Identifier of the conversion event
client_conversion_event_local_created_at	TIMESTAMP	Time when the client conversion event was created at (Local Time).
client_conversion_event_name	VARCHAR	Name of the client conversion event
client_conversion_event_source	VARCHAR	Source of the client conversion event

Col Name	Type	Description
client_conversion_event_filter	VARCHAR	Filterable label of the event which triggered the conversion for a user. (MEMBERSHIP, USER_CREDIT, ALL)

fct_client_conversion_events_part_2 (figure 1.6.3)

Dataset containing data which can be used as input to Part 2 task in the situation where no solution for Part 1 has been achieved.

Col Name	Type	Description
user_id	VARCHAR	Identifier for the user.
branch_id	VARCHAR	Identifier for the branch.
local_user_created_at	TIMESTAMP	Time when the user was created at (Local Time).
lead_status	VARCHAR	Lead status of the user either LEAD or CLIENT
client_conversion_event_type	VARCHAR	Event type of the conversion either MEMBERSHIP or USER_CREDIT
client_conversion_event_id	VARCHAR	Identifier of the conversion event
client_conversion_event_local_created_at	TIMESTAMP	Time when the client conversion event was created at (Local Time).
client_conversion_event_name	VARCHAR	Name of the client conversion event
client_conversion_event_source	VARCHAR	Source of the client conversion event
first_user_membership_id	VARCHAR	Identifier of the purchased membership associated with the first event found for a given user.
first_local_membership_purchased_at	TIMESTAMP	Time when the membership was purchased at (Local Time) associated with the first event found for a given user.
first_membership_name	VARCHAR	Membership name associated with the first event found for a given user.
first_membership_source	VARCHAR	Membership source associated with the first event found for a given user.
first_credit_pack_id	VARCHAR	Identifier of purchased credit pack associated with the first event found for a given user.

Col Name	Type	Description
first_local_credit_pack_purchase_at	TIMESTAMP	Time when the credit pack was purchased at (Local Time) associated with the first event found for a given user.
first_credit_pack_name	VARCHAR	Credit pack name associated with the first event found for a given user.
first_credit_pack_source	VARCHAR	Credit Pack source associated with the first event found for a given user.