Table of Contents

[**I. Question A** 1](#_heading=h.gjdgxs)

[**II. Question B** 2](#_heading=h.30j0zll)

[a. Import 2](#_heading=h.1fob9te)

[b. Metadata 2](#_heading=h.3znysh7)

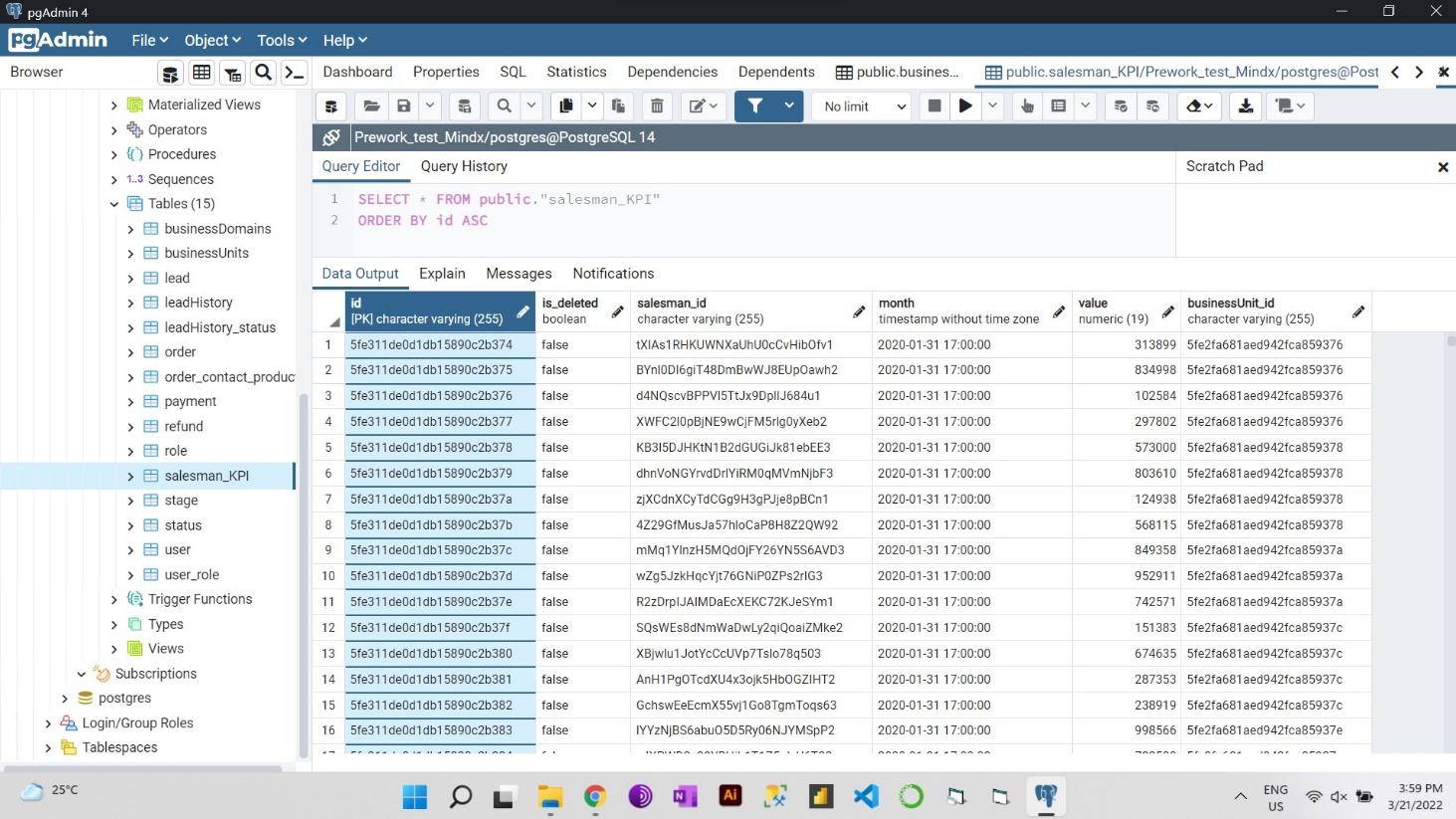
[c. Relationship 4](#_heading=h.2et92p0)

[d. The stages of a lead 6](#_heading=h.tyjcwt)

# I. Question A

*(Download the test data and restore it into a PostgreSQL database)*

* Sau khi tải file dữ liệu em tiến hành restore dữ liệu vào PostgreSQL. Ta tạo một Database mới để đưa dữ liệu vào.
* Sau đó em đưa dữ liệu từ file SQL vào database bằng CMD (em làm theo link hướng dẫn của HR gửi Link). Tuy nhiên vẫn ko restore được nên em search lỗi. Sau đó cài đặt lại trong Control Panel > System Properties > Advanced > Enviroment Variables và restore thành công.



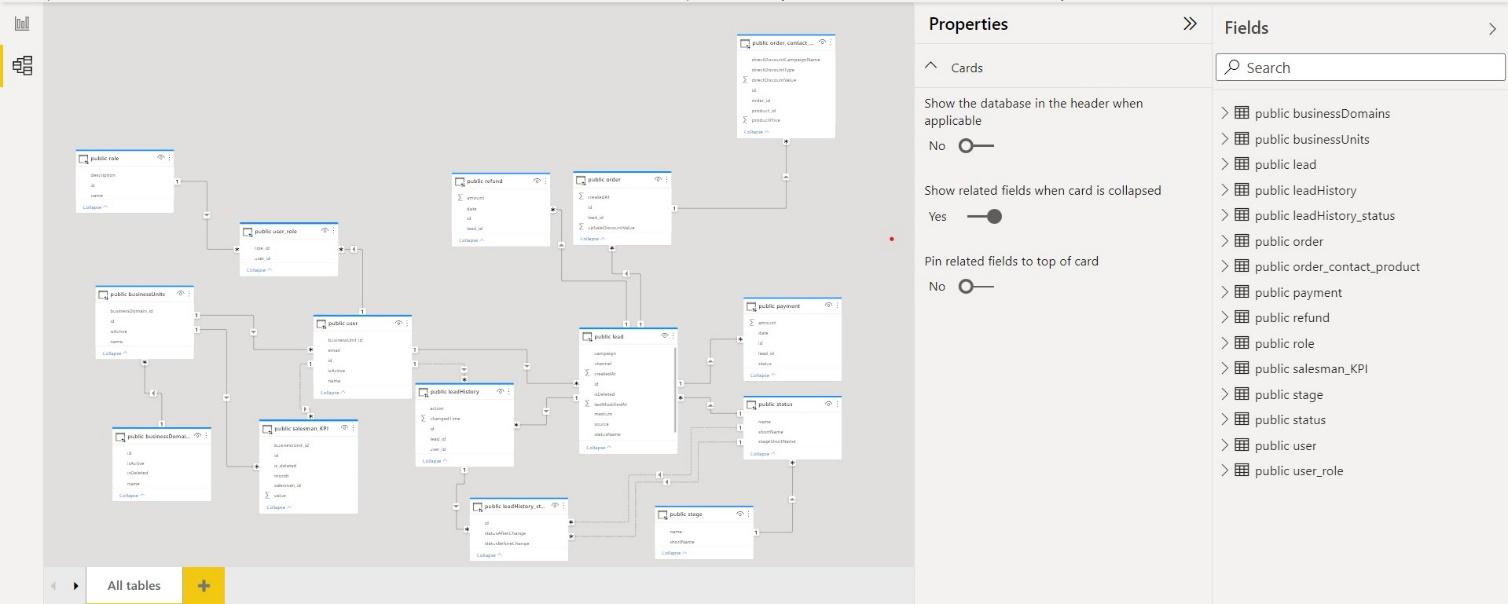
# II. Question B

*Import the given data into Power BI then have look at the data model holistically, and get a general idea of what the dataset is describing. Answer the following questions:*

* *How many tables are there in the dataset? Can you make a guess of what each table describes?*
* *What are the relationships between the tables? What can you tell from these relationships?*
* *What are the stages of a lead (of customer) in the system?*

## a. Import

Import dữ liệu từ PortgreSQL vào PowerBi (Data tab > Get data > PortgreSQL Database)



## b. Metadata

*(How many tables are there in the dataset? Can you make a guess of what each table describes?)*

**Ta có tổng cộng 15 bảng trong dataset**

***Bảng 0-1: Metadata***

| **No** | **Table** | **Describe** | **Note** |
| --- | --- | --- | --- |
| 1 | businessDomains | Những lĩnh vực kinh doanh chính của Cty | XSchool, Iliat |
| 2 | businessUnits | Các đơn vị kinh doanh | 23 đơn vị |
| 3 | lead | Danh mục khác hàng tiềm năng | Cá nhân, tổ chức có để lại thông tin |
| 4 | leadHistory | Lịch sử hoạt động của Lead | 19 hoạt động |
| 5 | leadHistory\_status | Lịch sử trạng thái của Lead | Dựa theo 2 bảng Stage và Status |
| 6 | order | Danh sách đơn hàng của lead | Lưu thông tin ngày tạo đơn của các Lead |
| 7 | order\_contact\_product | Thông tin đơn hàng | Lưu thông tin về giá cả và giảm giá của các đơn hàng |
| 8 | payment | Các khoản thanh toán | Thông tin về các đơn hàng đã được thanh toán (ngày giờ, số tiền) |
| 9 | refund | Các khoản thanh toán bị trả lại | các đơn hàng đã bị trả lại (ngày giờ, số tiền) |
| 10 | role | Các chức vụ trong Cty | 23 vị trí |
| 11 | salesman\_KPI | KPI của sale | Thông tin về doanh số của sale trong tháng |
| 12 | stage | Các giai đoạn của một Lead | Từ L0 đến L5 |
| 13 | status | Trạng thái của Lead | Khi Cty tương tác với khách hàng |
| 14 | user | Danh sách người dùng trong hệ thống | Thông tin người dùng, còn hoạt động hay không |
| 15 | user\_role | Chức vụ User | Cho biết chức vụ của user trong hệ thống |

## c. Relationship

*(What are the relationships between the tables? What can you tell from these relationships?)*

Mối quan hệ giữa các bảng trong bộ dataset là One to Many và One to One. Bộ dữ liệu trong dataset tập trung chủ yếu xoay quanh giữa 2 bảng fact là Lead và User liên kết với các bảng dimension theo mô hình **Snow Schema**. Mối quan hệ của các bảng được nêu trong bảng dưới đây:

***Bảng 0-2: Relationship***

| **No** | **Table** | **Relationship (Table number)** | **Describe** |
| --- | --- | --- | --- |
| 1 | businessDomains | 2 | Cung cấp thông tin chi tiết về lĩnh vực kinh doanh |
| 2 | businessUnits | 1,11,14 | Thông tin về lĩnh vực hoạt động của saleman và user |
| 3 | lead | 14,13,4,6,8,9 | Bảng Fact - Thông tin chi tiết về các nguồn của khách hàng tiền năng |
| 4 | leadHistory | 14,3,5 | lịch sử tương tác của các lead |
| 5 | leadHistory\_status | 4,13 | Trạng thái của các lead trước và sau khi tương tác |
| 6 | order | 3,7 | Cho biết ngày tạo đơn hàng của các lead |
| 7 | order\_contact\_product | 6 | Thông tin khoản thanh toán và giảm giá của sản phẩm |
| 8 | payment | 3 | Thông tin thanh toán của lead |
| 9 | refund | 3 | Thông tin trả lại của lead |
| 10 | role | 15 | Vị trí của các User |
| 11 | salesman\_KPI | 2,14 | Danh sách doanh thu của saleman qua đơn vị kinh doanh và chức vụ của saleman |
| 12 | stage | 13 | Trạng thái của Stage |
| 13 | status | 3,5,12 | Thông tin về Stage trước và sau tương tác |
| 14 | user | 2,3,4,11,15 | Cho biết user nào tương tác với các lead |
| 15 | user\_role | 10, 14 | Thông tin về vị chí người dùng qua id |

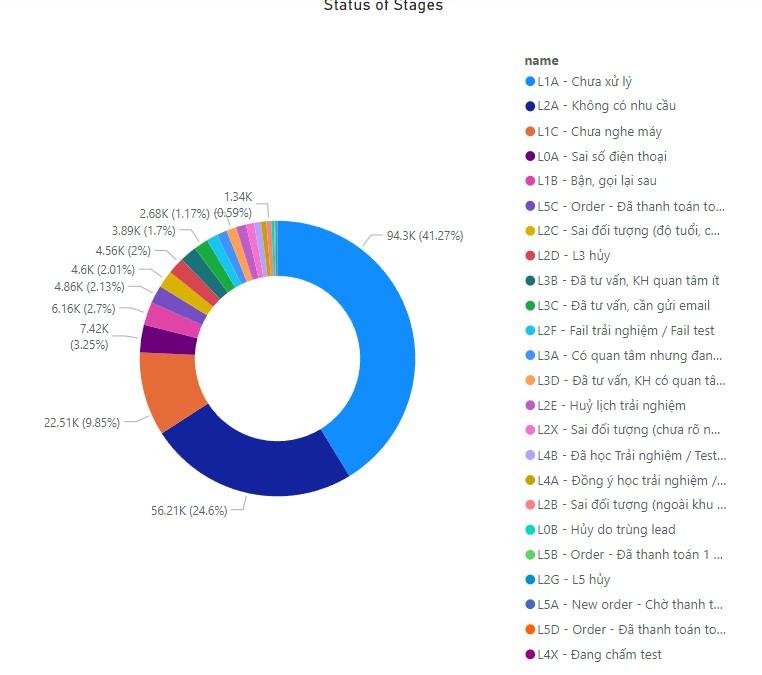
## d. The stages of a lead

*(What are the stages of a lead (of customer) in the system?)*

**Bộ dataset cho ta thấy có 6 cấp độ lead (khách hàng tiềm năng) từ L0 đến L5.**

Mỗi cấp độ Lead là một trạng thái của một khách hàng tiềm năng từ khi mới tiếp xúc với công ty cho đến trở thành khách hàng chính thức.

* **L0 - Junc Lead**: là cấp độ thấp nhất của Lead có để lại thông tin nhưng saleman chưa thể tương tác .Với 2 loại là :
  + **L0A** (Sai số điện thoại)
  + **L0B** (Hủy do trùng Lead)
* **l1 - Lead**: là các Lead chưa được saleman xử lí hoặc đã liên hệ nhưng liên lạc lại. Bao gồm 3 loại:
  + **L1A** (Chưa xử lý),
  + **L1B** (Bận, gọi lại sau),
  + **L1C** (Chưa nghe máy).
* **L2 - Cold Lead**: là các lead đã liên lạc theo thông tin đã có nhưng từ chối sử dụng sản phẩm dịch vụ vì nhiều lí do. Bao gồm 8 loại:
  + **L2A** - Không có nhu cầu
  + **L2B** - Sai đối tượng (ngoài khu vực)
  + **L2C** - Sai đối tượng (độ tuổi, chưa có con...)
  + **L2D** - L3 hủy
  + **L2E** - Huỷ lịch trải nghiệm
  + **L2F** - Fail trải nghiệm / Fail test
  + **L2G** - L5 hủy
  + **L2X** - Sai đối tượng (chưa rõ nguyên nhân)
* **L3 - Warm Lead**: Là các Lead đã được saleman liên lạc và quan tâm. Bao gồm 4 loại:
  + **L3A** - Có quan tâm nhưng đang bận,
  + **L3B** - Đã tư vấn, KH quan tâm ít,
  + **L3C** - Đã tư vấn, cần gửi email,
  + **L3D** - Đã tư vấn, KH có quan tâm nhiều.
* **L4 - Hot Lead**: là các đối tượng Lead quan tâm đến sản phẩm, dịch vụ. Bao gồm 3 loại:
  + **L4A** - Đồng ý học trải nghiệm / Test,
  + **L4B** - Đã học Trải nghiệm / Tested,
  + **L4X** - Đang chấm test.
* **L5 - Order**: Là khách hàng đã mua và thanh toán hoặc chờ thanh toán sản phẩm.
  + **L5A** - New order - Chờ thanh toán,
  + **L5B** - Order - Đã thanh toán 1 phần,
  + **L5C** - Order - Đã thanh toán toàn phần,
  + **L5D** - Order - Đã thanh toán toàn phần nhưng không học.



* Các Lead ban đầu được lấy thông tin từ nhiều kênh như social network, website, ads, các sự kiện, email,... Sau đó được chuyển qua cho các User để xử lý. Dựa trên sự tương tác với công ty người dùng đã ở trong **sales funnel** (the lead stage).
* Dữ liệu có trong bộ dataset đều từ trong mô hình **CRM** - Customer relationship management của công ty.
* Lead là danh sách các khác hàng tiềm năng tuy nhiên không phải tất cả các khách hàng tiềm năng đều sẽ mua hàng.
* Các lead trong giai đoạn L5 có khả năng refund sản phẩm/dịch vụ khi đã thanh toán là tương đối nhiều.



***Xin chân thành cảm ơn !***