

**Họ Tên: Mai Anh Tuấn**

Số điện thoại : +84797996113

Email : [dannymai2012@gmail.com](mailto:dannymai2012@gmail.com)

Ngày tham gia : 22/11/2022

### **Câu hỏi số 6:**

Đầu tiên, tôi chia project này thành ba đầu mục lớn:

- Chuẩn bị dữ liệu
- Xử lý dữ liệu
- Trực quan hóa dữ liệu

#### **I. Chuẩn bị dữ liệu**

##### **1. Kiểm tra dữ liệu đầu vào**

Các đặc điểm của dữ liệu:

- Header name : Đầy đủ thông tin, không thiếu sót, không trùng lặp
- Null missing : Bị ở ba cột là staff\_id , cost/GSV , category
- Outlier : Bị ở hầu hết các cột numeric tuy nhiên sẽ không xử lý chi tiết outlier vì chưa thật sự hiểu sâu về domain business
- Blank row : Không có
- Wrong format : Không có

##### **2. Lập User Story cho dữ liệu**

Thực hiện lập User Guide cho dữ liệu để khái quát thông tin dữ liệu và mô phỏng quá trình lên ý tưởng phác thảo cho khâu trực quan hóa ở các bước sau.

USER STORY ID	PRIORITY	AS A [type of user]	I NEED TO [do some task]	SO THAT I CAN [get some result]	STATUS
1	Must have	As all people	To know about this dashboard	Information Board	Approved
2	Must have	As a Director/Sales manager	To get overview perspective	Sales Overview	Approved
3	Must have	As a sales Manager	To know Sales Performace, Territory, Customer Report	Sales Performance, Territory, Customer Report	Approved

## II. Xử lý dữ liệu

Sử dụng Google Sheet và Power BI để xử lý dữ liệu.

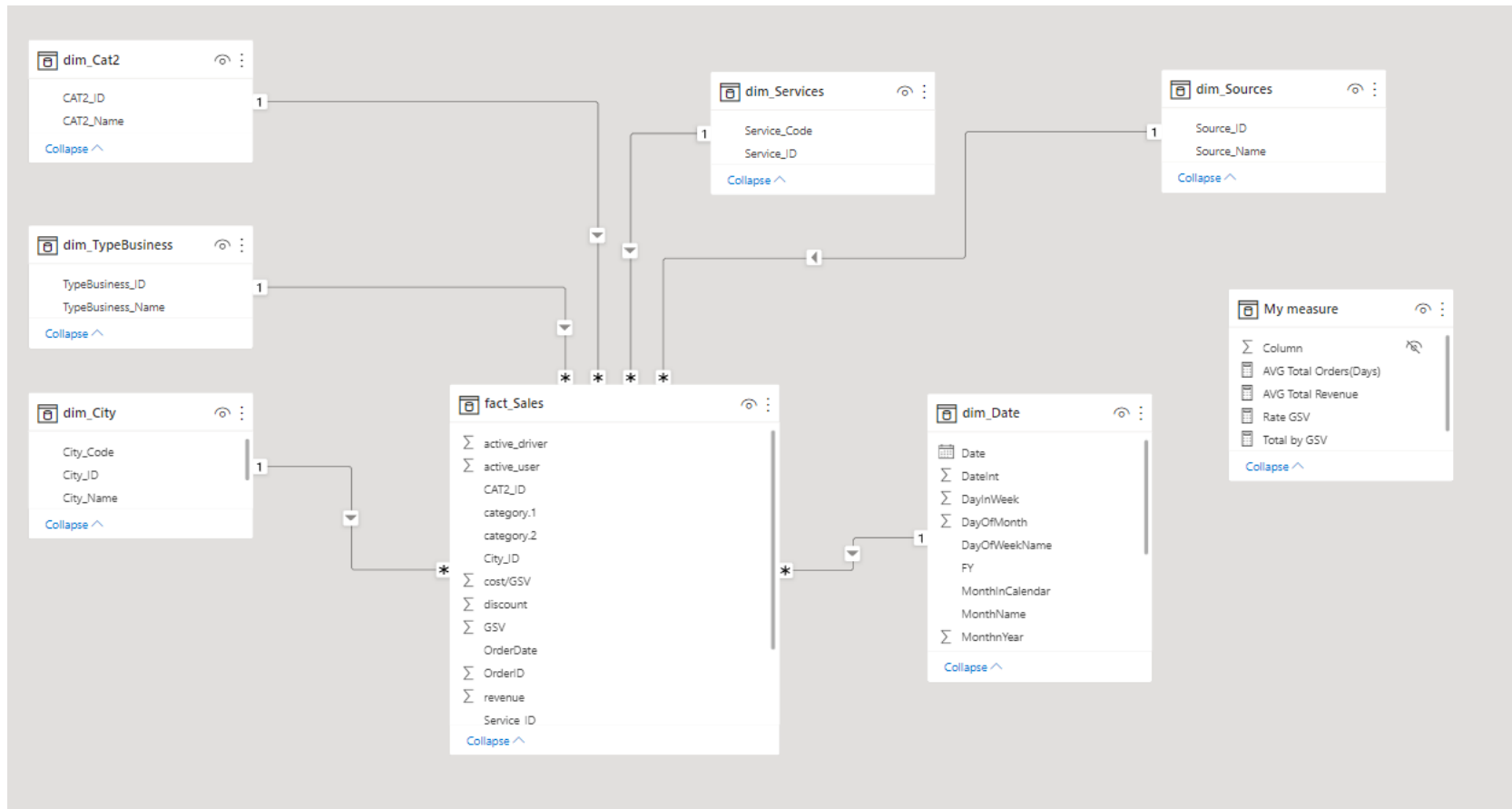
Đầu tiên xử lý các Null Missing của dữ liệu trong các cột staff\_id , cost/GSV , category. Đối với cột category các giá trị null được thay thế bằng Unknown hoặc undefined. Các giá trị null missing trong cột staff\_id được xử lý thay thế bằng số 0. Cột cost/GSV không xử lý vì dữ liệu đó sẽ được thay thế bằng các metric khác thực hiện bằng hàm tính toán trong Power BI.

Thực hiện một số thao tác xử lý khác như sửa tên cột cho trực quan hơn, thay thế định dạng các cột ngày tháng, chuyển data format từ decimal sang currency đối với các cột có đại lượng tính toán là tiền tệ như cột Revenue, cột GSV,.v...v....

Thực hiện tách các dữ liệu thành các bảng dimension\_ và fact\_Sales để tạo model hoàn chỉnh phục vụ cho giai đoạn trực quan hóa dữ liệu. Các bảng được tạo thêm là dim\_Category, dim\_City, dim\_Services, dim\_Source, dim\_TypeBusiness, dim\_Cat2. Viết function code để tạo bảng dim\_Date tự động dùng để sử dụng trong quá trình trực quan hóa dữ liệu ở bước sau. Tạo thêm các cột index trong các bảng dimension và bảng fact\_Sales để nối các bảng theo mô hình Star Schema hoàn chỉnh. Trong bảng dim\_City bổ sung thêm 2 cột Longitude và Latitude để trực quan hóa bằng bản đồ trong Dashboard tại tab Territory.

Tạo thêm tổng cộng 5 metrics bao gồm

- AVG Total Orders (Days): tính trung bình tổng số đơn trong một ngày
- AVG Total Revenue : tính trung bình tổng doanh thu
- Rate GSV : tính tỷ lệ % giữa chi phí là doanh thu thuần theo công thức  $\text{cost}(\text{discount}) / \text{GSV}$
- Total by GSV : tính tổng giá trị GSV
- Total orders : tính tổng số đơn hàng



### III. Trục quan hóa dữ liệu

Xem tại link:

<https://app.powerbi.com/view?r=eyJrIjoibNzZMGjYjEtNjhkNy00MWUxLTljNDMtMzFhNGFhNGRjMDFmliwidCI6IjdhODAwMTEwLWJmNGEtNDkwOC1hZml4LTdjNWVjNzNhMDE0MyIsImMiOiJlEwfQ%3D%3D>