

**CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**  
**Độc lập - Tự do- Hạnh phúc**

**BÁO CÁO HỌC TẬP NHÓM**

*Kính gửi:* Giảng viên bộ môn Lập Trình Python nâng cao- thầy Cao Diệp Thắng

*Tên nhóm:* nhóm 3

Lớp: DHKL18A3HN

*Thành viên nhóm:*

- Đỗ Thị Vân Anh- 24174600128
- Đinh Diệu Linh- 24174600172
- Nguyễn Thùy Linh- 24174600134
- Hoàng Thị Mai- 24174600153
- Hoàng Nguyễn Khánh Vy- 24174600156

*Tên chủ đề :* CASE STUDY 13 – DỮ LIỆU TIÊM CHỦNG & SỨC KHỎE CỘNG ĐỒNG

*Thời gian hoàn thành:* trước 0 giờ ngày 17/12/2025.

**I. Giới thiệu chủ đề:**

Dữ liệu tiêm chủng thường được thu thập từ nhiều nguồn như trạm y tế, bệnh viện, chiến dịch tiêm chủng lưu động. Điều này dẫn đến dữ liệu không đồng nhất, dễ phát sinh lỗi về mã số bệnh nhân, hình thức ghi ngày tháng, cách viết tên vaccine hoặc mức độ phản ứng sau tiêm.

Case study này giúp rèn luyện kỹ năng khảo sát – làm sạch – tổng hợp – phân tích dữ liệu y tế thực tế theo đúng quy trình chuẩn.

**II. Phương pháp xử lí và phân tích dữ liệu**

Tên thành viên	Nhiệm vụ	Kỹ thuật sử dụng	Hoàn thành
Nguyễn Thùy Linh	<b>Task 1 – Đọc &amp; làm sạch dữ liệu</b> - Khảo sát cấu trúc 3 bảng - Chuẩn hoá mã và nhãn quan trọng. - Xử lý dữ liệu thiếu và giá trị bất thường. - Kiểm tra tính hợp lệ logic - Đảm bảo khả năng merge - Xuất các file sạch	read_csv(), to_csv(), dropna(), fillna(), astype(), to_datetime(), str.strip(), str.title()	x
Hoàng Nguyễn Khánh Vy	<b>Task 2 – Truy vấn &amp; thống kê mô tả:</b> - Đếm số bệnh nhân theo nhóm tuổi. - Tìm bệnh nhân đã tiêm từ 2 loại vaccine trở lên. - Thống kê số lần tiêm theo loại vaccine. - Liệt kê các bản ghi có dose_number bất thường	loc, iloc, isin	x
Hoàng Thị Mai	<b>Task 3 – GroupBy &amp; Tổng hợp:</b> - Tổng số liều tiêm theo nhóm tuổi và loại vaccine. - Trung bình số liều tiêm theo giới tính. - Tỷ lệ bệnh nhân có phản ứng sau tiêm theo vaccine_type.	groupby(), agg()	x

	<ul style="list-style-type: none"> <li>Mức độ phản ứng trung bình theo nhóm tuổi.</li> </ul>		
Đinh Diệu Linh	<p><b>Task 4 – Merge dữ liệu:</b></p> <ul style="list-style-type: none"> <li>Merge patient_info với vaccination_record theo patient_id.</li> <li>Merge thêm reaction_log</li> <li>Phát hiện dữ liệu bị thiếu</li> <li>Tạo bảng full_outer_merge để liệt kê tất cả bệnh nhân và phân tích nguyên nhân lệch dữ liệu.</li> </ul>	merge() với inner/left,	x
Đỗ Thị Vân Anh	<p><b>Task 5 – Pivot Table + Stack/Unstack:</b></p> <ul style="list-style-type: none"> <li>Tạo pivot tổng số liều tiêm theo nhóm tuổi × vaccine_type.</li> <li>Tạo pivot mức độ phản ứng trung bình theo vaccine_type × age_group.</li> <li>Thực hành stack/unstack</li> <li>Tạo pivot đa cấp để thống kê số bệnh nhân, số lần tiêm và tỷ lệ có phản ứng</li> </ul>	pivot_table(), "stack(), unstack()	x

### III. Kết quả :

Bảng 1: THỐNG KÊ SỐ LẦN TIÊM THEO LOẠI VẮC XIN

vaccine_type    số lần tiêm		
0	Covid-19	48
1	DPT	36
2	MMR	36

Nhận xét:

- Vaccine Covid-19 có số lần tiêm cao nhất, khoảng 48 lần, cho thấy đây là loại vaccine được sử dụng nhiều nhất.
- Hai loại DPT và MMR có số lần tiêm tương đương nhau (khoảng 36 lần), thấp hơn rõ rệt so với Covid-19.
- Nhìn chung, Covid-19 chiếm tỷ trọng lớn nhất, còn DPT và MMR đóng vai trò bổ sung với mức độ tiêm tương đối ổn định.

Bảng 2: Bảng liệt kê tất cả bệnh nhân:

patient_id	age_group	gender	vaccine_type	dose_number	vaccination_date	merge_patient.vaccine	reaction_type	severity	report_date	vaccination_date_min
0	BN500	6-17	F	Covid-19	2.0	2024-04-15	both	Pain	3.0	NaN
1	BN500	6-17	F	MMR	1.0	2024-04-04	both	Pain	3.0	NaN
2	BN500	6-17	F	Covid-19	2.0	NaN	both	Pain	3.0	NaN
3	BN500	6-17	F	DPT	1.0	NaN	both	Pain	3.0	NaN
4	BN500	6-17	F	DPT	NaN	2024-05-11	both	Pain	3.0	NaN

Nhận xét:

- Có 258 bệnh nhân có đầy đủ dữ liệu tiêm và phản ứng, chiếm đa số trong tập dữ liệu.
- Còn tồn tại 17 trường hợp có dữ liệu tiêm thiếu phản ứng
- Có 2 trường hợp có phản ứng nhưng không có dữ liệu tiêm

Bảng 3: Thống kê số bệnh nhân, số lần tiêm và tỉ lệ có phản ứng

Bảng pivot đa cấp để thống kê số bệnh nhân, số lần tiêm và tỷ lệ có phản ứng:									
	patients				doses				reaction_rate \ 0-5
age_group	0-5	6-17	18-49	50+	0-5	6-17	18-49	50+	
vaccine_type									
Covid-19	3	9	8	8	5	30	36	27	100.0
DPT	3	5	7	7	9	12	29	42	100.0
MMR	4	7	7	4	13	21	36	15	100.0

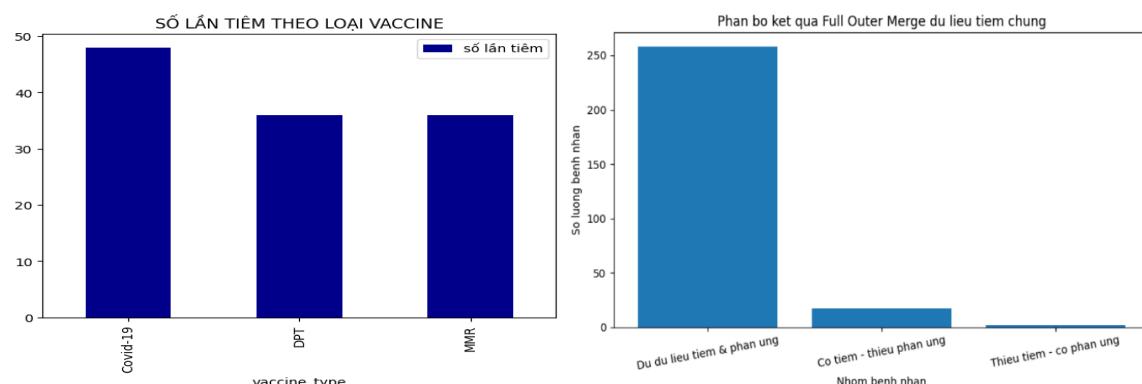
  

age_group	6-17	18-49	50+
vaccine_type			
Covid-19	100.0	100.0	100.0
DPT	100.0	100.0	100.0
MMR	100.0	100.0	100.0

Nhận xét:

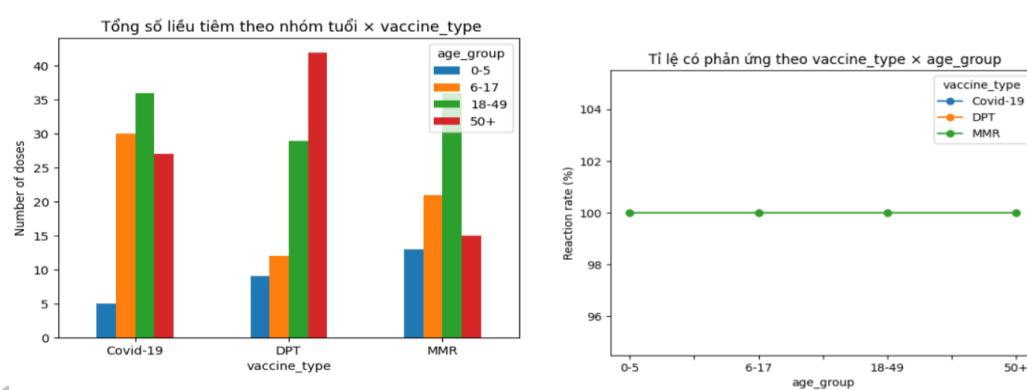
- Số bệnh nhân và số lần tiêm phân bố khá đồng đều giữa các nhóm tuổi, trong đó nhóm 18–49 có số liều tiêm cao nhất.
- Tỷ lệ ghi nhận phản ứng sau tiêm ở tất cả vaccine và nhóm tuổi đều đạt 100%, cho thấy dữ liệu phản ứng được ghi nhận đầy đủ

#### IV. Biểu đồ:



*Biểu đồ 1 (Task 2):  
Số lần tiêm theo loại vaccine*

*Biểu đồ 2 (Task 4):  
Phân bố kết quả dữ liệu tiêm chung*



*Biểu đồ 3 (Task 5):  
Tổng số liều tiêm theo nhóm tuổi và loại vaccine*

*Biểu đồ 4 (Task 5):  
Tỉ lệ có phản ứng theo nhóm tuổi và loại vaccine*

## V. Nhận xét & Đánh giá:

- Vaccine Covid19 có số lượng tiêm nhiều nhất, độ an toàn cao (phản ứng nhẹ).
- Nhóm 18–49 là đối tượng tiêm chủ yếu trong cộng đồng. Các nhóm tuổi lớn hơn có xu hướng phản ứng mạnh hơn sau tiêm.
- Một phần dữ liệu thiếu (vaccine hoặc phản ứng) cho thấy cần cải thiện quy trình nhập liệu từ các cơ sở y tế

## VI. Kết luận:

Qua case study này, nhóm đã:

- Hiểu rõ quy trình xử lý dữ liệu y tế từ dữ liệu thô đến dữ liệu phân tích.
- Rèn luyện kỹ năng làm sạch, chuẩn hóa và kiểm tra logic dữ liệu.
- Vận dụng thành thạo các thao tác GroupBy, Merge, Pivot Table trong Pandas.
- Rút ra được các thống kê quan trọng phục vụ đánh giá chương trình tiêm chủng.
- Kết quả phân tích có thể được sử dụng làm cơ sở cho các báo cáo y tế, hỗ trợ ra quyết định và xây dựng chính sách tiêm chủng hiệu quả hơn trong thực tế.

*Nhóm 3 xin chân thành cảm ơn!*

**NGƯỜI BÁO CÁO**

*(Ký và ghi rõ họ tên)*

Vân Anh

Đỗ Thị Vân Anh