Câu 3:

1. **Bước 1:**
2. Import thư viện
3. Mối liên hệ giữa các thuộc tính:

* Lập heatmap tìm ra các thuộc tính ảnh hưởng nhiều đến Stroke.
* Sau đó chọn ra những thuộc tính có tương quan cao: age, hypertension, hear\_disease, bmi, ever\_married.

1. **Bước 2:** Gom những thuộc tính trên thành các nhóm 2,3 để vẽ biểu đồ xem xét ảnh hưởng của chúng đối với stroke.
2. Giới tính và tình trạng hút thuốc
3. Tiền sử bệnh tim và cao huyết áp
4. Cao huyết áp, độ tuổi trung bình, tình trạng hút thuốc
5. BMI và loại công việc
6. BMI và giới tính
7. BMI và tuổi tác và loại công việc
8. BMI và loại công việc và giới tính

Câu 4: Mô hình đột quỵ

* Code 1,2: Xử lý các giá trị null
* Code 3: Tiền xử lý, Làm phẳng các lớp bằng kỹ thuật SMOTE
* Code 4,5:

+ Cài đặt model, đồng thời truyền tham số cho mỗi model bằng phương pháp hyperparameter và GridSearhCV để tăng độ chính xác.

+ Lý do phải chia thành 2 phần là vì 2 phương pháp truyền tham số trên tốn nhiều không gian bộ nhớ, có nguy cơ dẫn đến dead cell !   
+ Những dòng được in ra ở model Lightgbm là output in ra, không phải lỗi !

* Code 6,7:

+ Vẽ biểu đồ trực quan hóa các thông số phản ánh hiệu suất của tất cả model

+ Từ đó ta chọn ra được model nào nên sử dụng để dự đoán cho tập dữ liệu có đặc trưng như này.