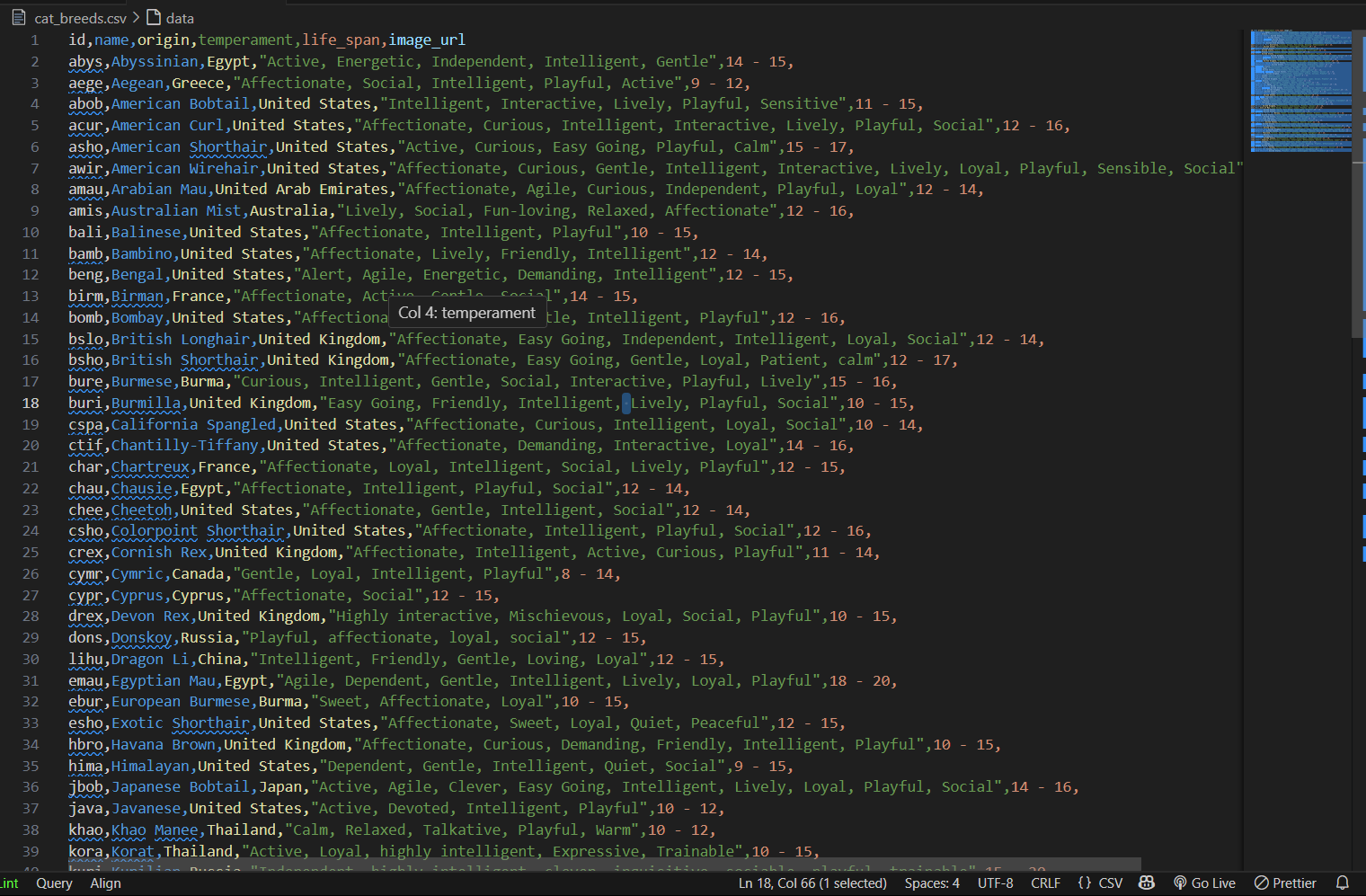
Họ và tên: Nguyễn Trần Phúc Khang

MSSV: 23660931

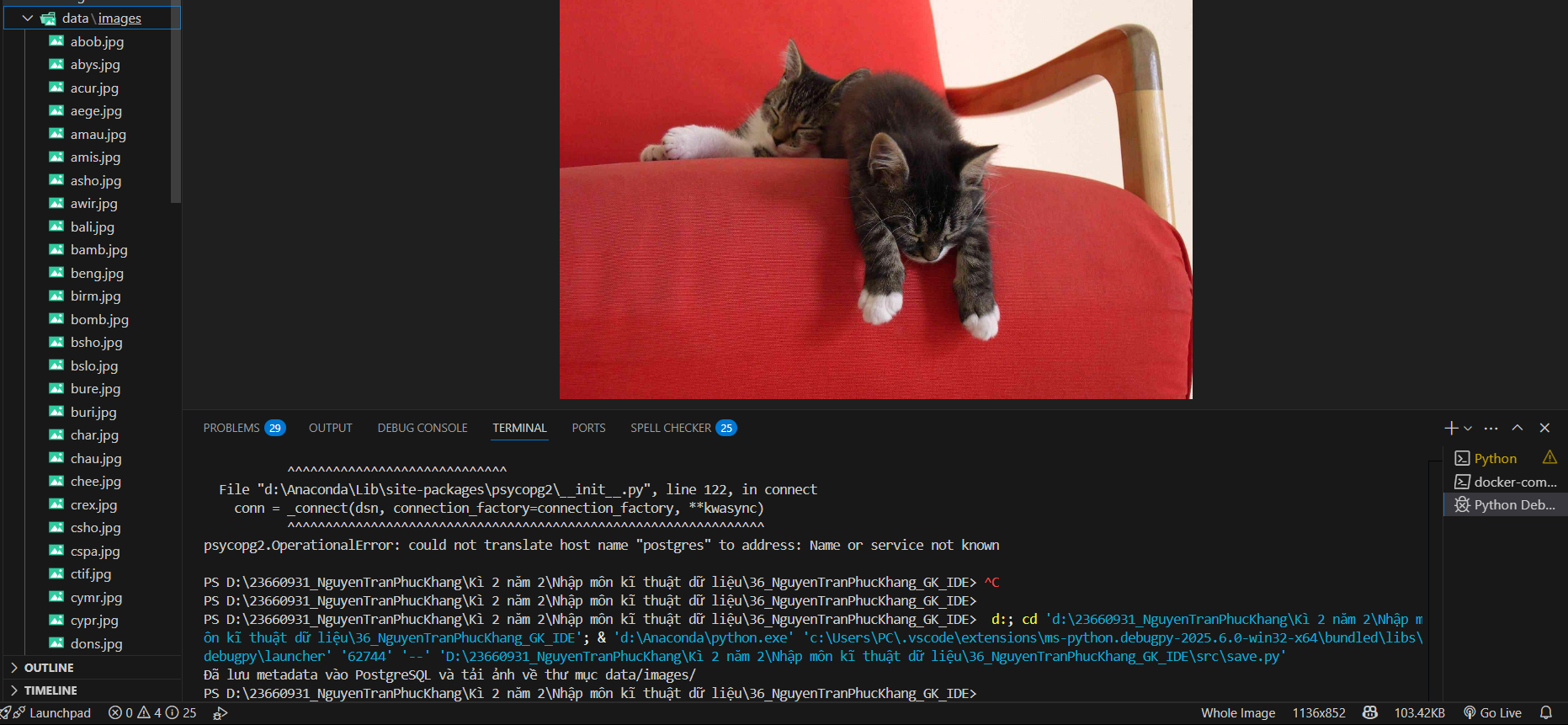
1. (2 điểm) Viết script (Python, curl, javascript,... or any language) để crawl thông tin về các giống mèo từ trang web https://api.thecatapi.com/v1/breeds , lưu trữ các thông tin sau: - - - - - - id: mã giống mèo name: tên giống mèo origin: nguồn gốc temperament: đặc tính life\_span: tuổi thọ image\_url: link ảnh (trường image.url)



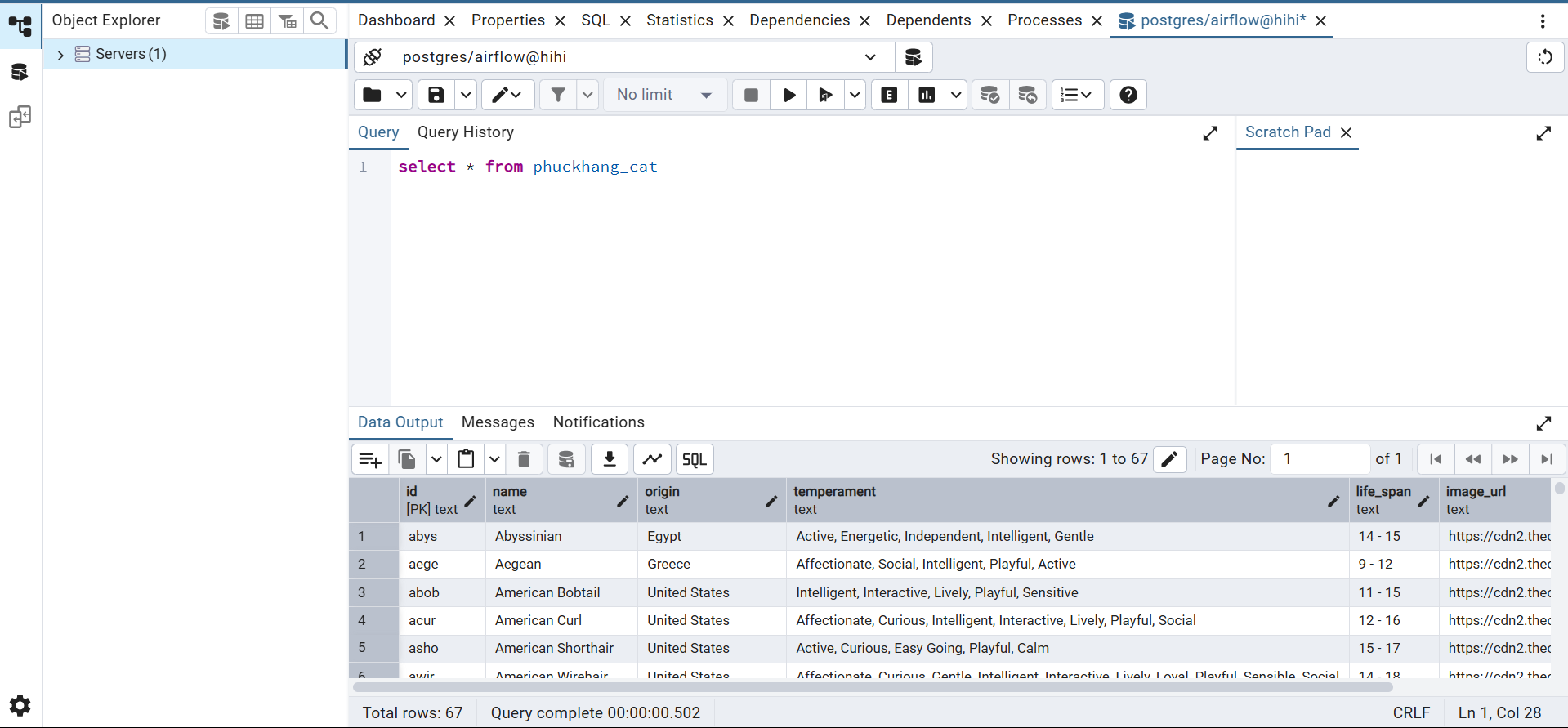
1. (2 điểm) Viết script python để làm sạch dữ liệu - Chuẩn hoá temperament (tính khí) thành dạng list (ví dụ: "Affectionate, Intelligent" → ["Affectionate", "Intelligent"]) - Nếu image\_url bị thiếu → Gán ảnh default: <https://cdn2.thecatapi.com/images/aae.jpg>



1. (2 điểm) Viết script python để lưu vào PostgreSQL (có script tạo bảng, kết nối DB) - Metadata: lưu vào PostgreSQL. - Hình ảnh: tải ảnh về thư mục local data/images/ theo định dạng .jpg.



1. (2 điểm) Viết Dockerfile + docker-compose.yml để đóng gói và chạy ứng dụng và chạy app và PostgreSQL



e. (2 điểm) Viết DAG Airflow để chạy pipeline - 3 task: crawl\_cat → transform\_cat → save\_cat - - - Lịch chạy: 9h sáng mỗi ngày. Thử lại 2 lần nếu bị thất bại. Bật Airflow, kiểm tra log, chụp kết quả DAG thành công.