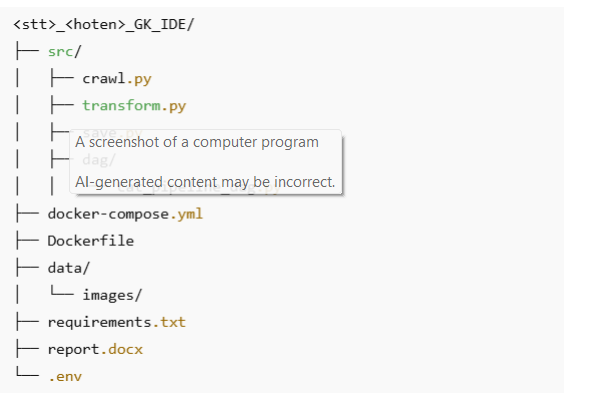
**Võ Khương Duy**

**MSSV: 23646941**

Đề bài: news\_pipeline ( 10 điểm)

Cấu trúc thư mục như sau:



Câu a. (2 điểm) Viết script python để Crawl

A screenshot of a computer program

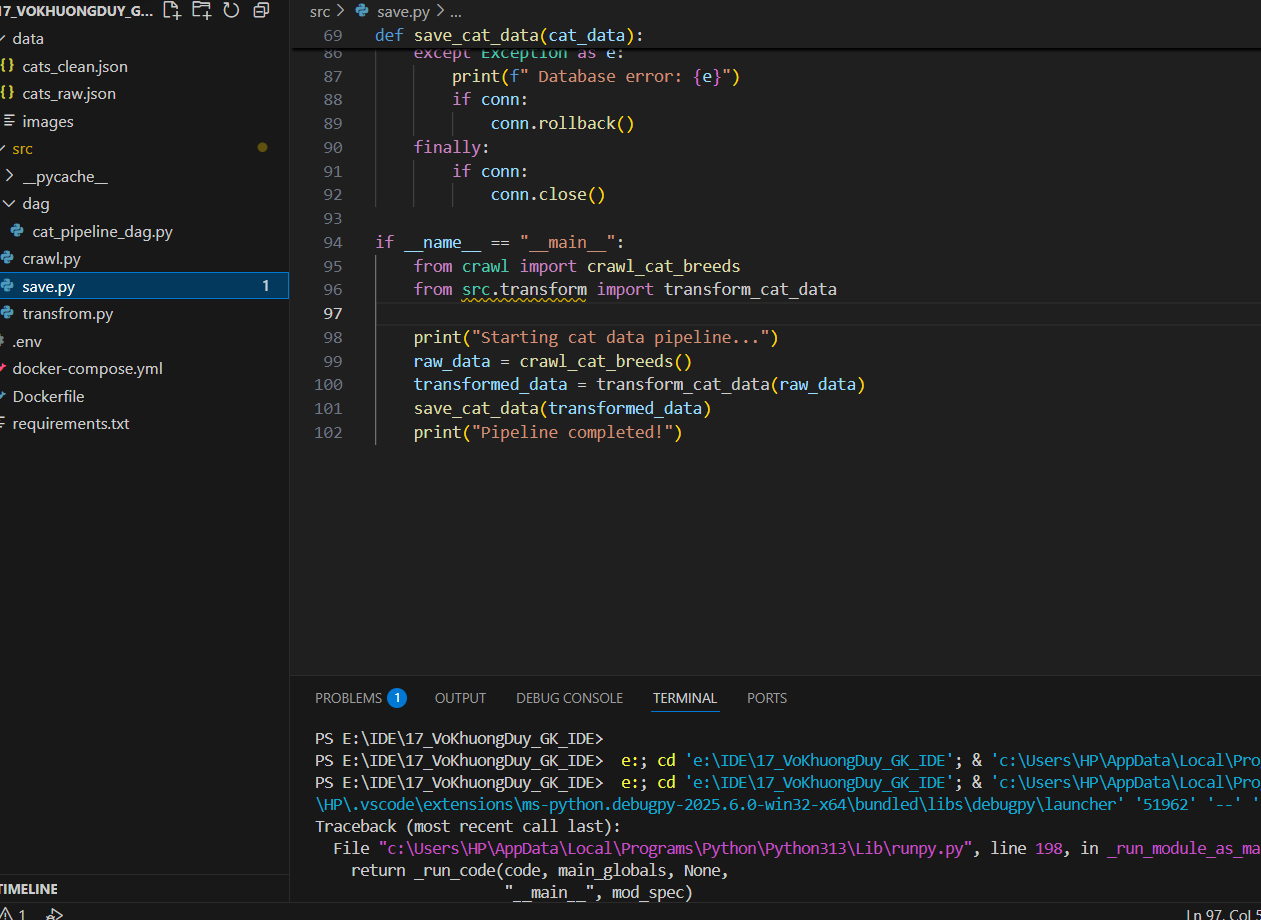
Description automatically generated

Câu b. (2 điểm) Viết script python để làm sạch dữ liệu:

A screenshot of a computer program

Description automatically generated

Câu c. (2 điểm) Viết script python để lưu vào PostgreSQL (có script tạo bảng, kết nối DB)



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

Câu e. (2 điểm) Viết DAG Airflow để chạy pipeline

A screen shot of a computer program

Description automatically generated

Chạy :docker-compose up –build

🡺Để Xây dựng lại hình ảnh Docker (image) và khởi động tất cả các dịch vụ

* Truy cập [DAGs - Airflow](http://localhost:8080/home)