1. Nguồn dữ liệu  
   Nguồn dữ liệu được sử dụng là Tiki - một trang web thương mại điện tử hàng đầu tại Việt Nam, cung cấp hàng ngàn sản phẩm đa dạng từ đồ điện tử, gia dụng, thời trang, sách, ...
2. Công cụ thu thập  
   Công cụ được sử dụng để thu thập dữ liệu là ngôn ngữ lập trình Python cùng với các thư viện: requests, pandas, csv, retrying. Công cụ này sử dụng API của Tiki để truy xuất và lấy dữ liệu từ trang web này thông qua các yêu cầu HTTP.
3. Các bước sử dụng công cụ  
   Các bước để thu thập dữ liệu từ Tiki qua API bao gồm:

* Tạo session để sử dụng cho tất cả các yêu cầu API của Tiki
* Đặt số lần retry mặc định là 10 để xử lý các lỗi liên quan đến mạng hoặc API phản hồi lỗi
* Xác định các headers, params, và body cần thiết cho từng yêu cầu API khác nhau
* Sử dụng hàm retrying để xử lý các lỗi xảy ra trong quá trình truy xuất API
* Tiến hành lấy dữ liệu bằng cách xác định các key chính của các đối tượng trả về, sau đó lưu trữ dữ liệu vào file CSV
* Thực hiện các bước trên cho tất cả các sản phẩm được hiển thị trên trang web Tiki qua các yêu cầu API khác nhau

1. Đầu vào và đầu ra của quá trình thu thập

* Đầu vào: Các yêu cầu API của Tiki được cấu trúc thông qua các headers, params, và body để truy xuất dữ liệu.
* Đầu ra: Các dữ liệu thu thập được sẽ được lưu trữ trên file CSV, các trường bao gồm id, tên sản phẩm, thương hiệu, giá gốc, giá giảm, giá hiện tại, tỷ lệ giảm giá, số lượng bán, đánh giá trung bình, số lượng đánh giá, mã nhà bán hàng, tên nhà bán hàng, ngày tham gia của nhà bán hàng, đánh giá trung bình của nhà bán hàng, số lượng đánh giá của nhà bán hàng và tổng số người theo dõi.

1. Ví dụ minh hoạ  
   Mã nguồn đã đưa ra ở câu trả lời trước cung cấp một ví dụ minh họa về cách sử dụng công cụ Python để lấy dữ liệu từ Tiki thông qua API. Ví dụ này đưa ra các bước trình bày trên để minh họa việc thu thập dữ liệu.

Trên cùng của Biểu mẫu



Dưới cùng của Biểu mẫu