**ĐẠI HỌC ĐÀ NẴNG  
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT**

Logo

Description automatically generated

**BÁO CÁO CUỐI KỲ  
CHUYÊN ĐỀ CƠ SỞ DỮ LIỆU**

**Tên đề tài:**

**CRAWL DATA WEBSITE  
BÁN LAPTOP**

**Giảng viên hướng dẫn : ThS. Nguyễn Thị Thúy Hoài**

**Sinh viên thực hiện : Nguyễn Thành Đạt**

**Mã sinh viên : 1811505310107**

**Lớp học phần : 220CCSDL03**

Đà Nẵng, tháng 7, năm 2021

**Mục lục**

[1. Tên đề tài 1](#_Toc76487242)

[2. Giải pháp công nghệ 1](#_Toc76487243)

[3. Link github file báo cáo 1](#_Toc76487244)

[4. Code 1](#_Toc76487245)

[5. Kết quả khi chạy 3](#_Toc76487246)

[5.1. Bảng sản phẩm 3](#_Toc76487247)

[6. Bảng danh mục cha 3](#_Toc76487248)

[6.1. Bảng danh mục con 4](#_Toc76487249)

[6.2. Hình ảnh sau khi được tải về máy 4](#_Toc76487250)

[7. Dữ liệu sau khi được xuất ra file exce 5](#_Toc76487251)

[7.1. Sheet Product 5](#_Toc76487252)

[7.2. Sheet Category 5](#_Toc76487253)

[7.3. Sheet Subcategory 6](#_Toc76487254)

[8. Sơ đồ ERD 6](#_Toc76487255)

## Tên đề tài

Craw data từ website philong.com lấy thông tin như tên sản phẩm, giá, thuộc danh mục, hình ảnh và xuất thông tin ra dưới dạng excel

Trang web dùng để lấy dữ liệu:

<https://philong.com.vn/may-tinh-xach-tay.html?filter=%2C118%2C>

## Giải pháp công nghệ

* Ngôn ngữ python phiên bản 3.9.5
* IDE sử dụng: Jupyter notebook
* Thư viện sử dụng: BeautifulSoup, urlopen, urllib.request, urlparse, os, pandas

## Link github file báo cáo

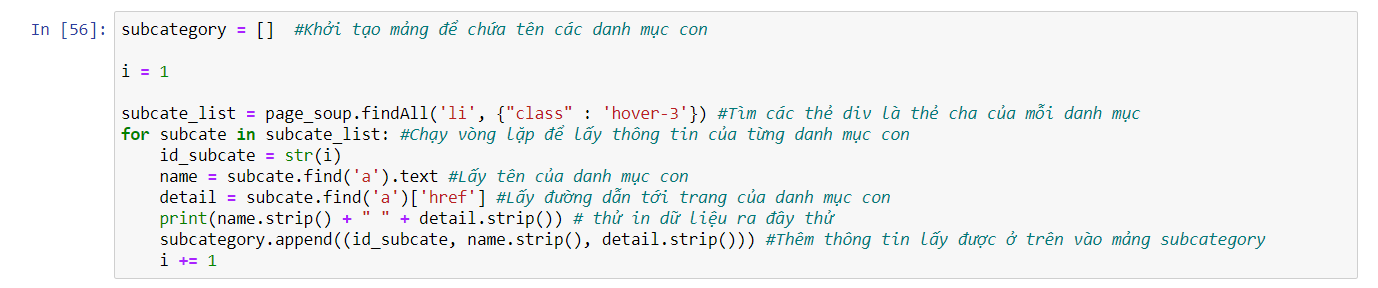
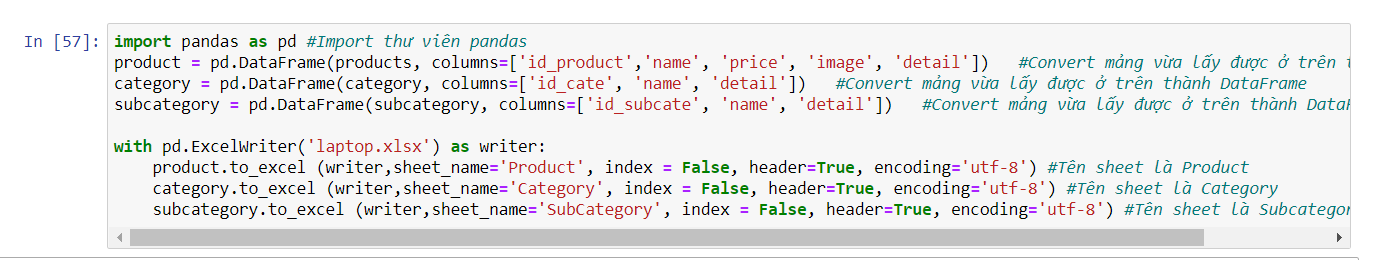
[datnguyennt/crawl-website (github.com)](https://github.com/datnguyennt/crawl-website/)

## Code



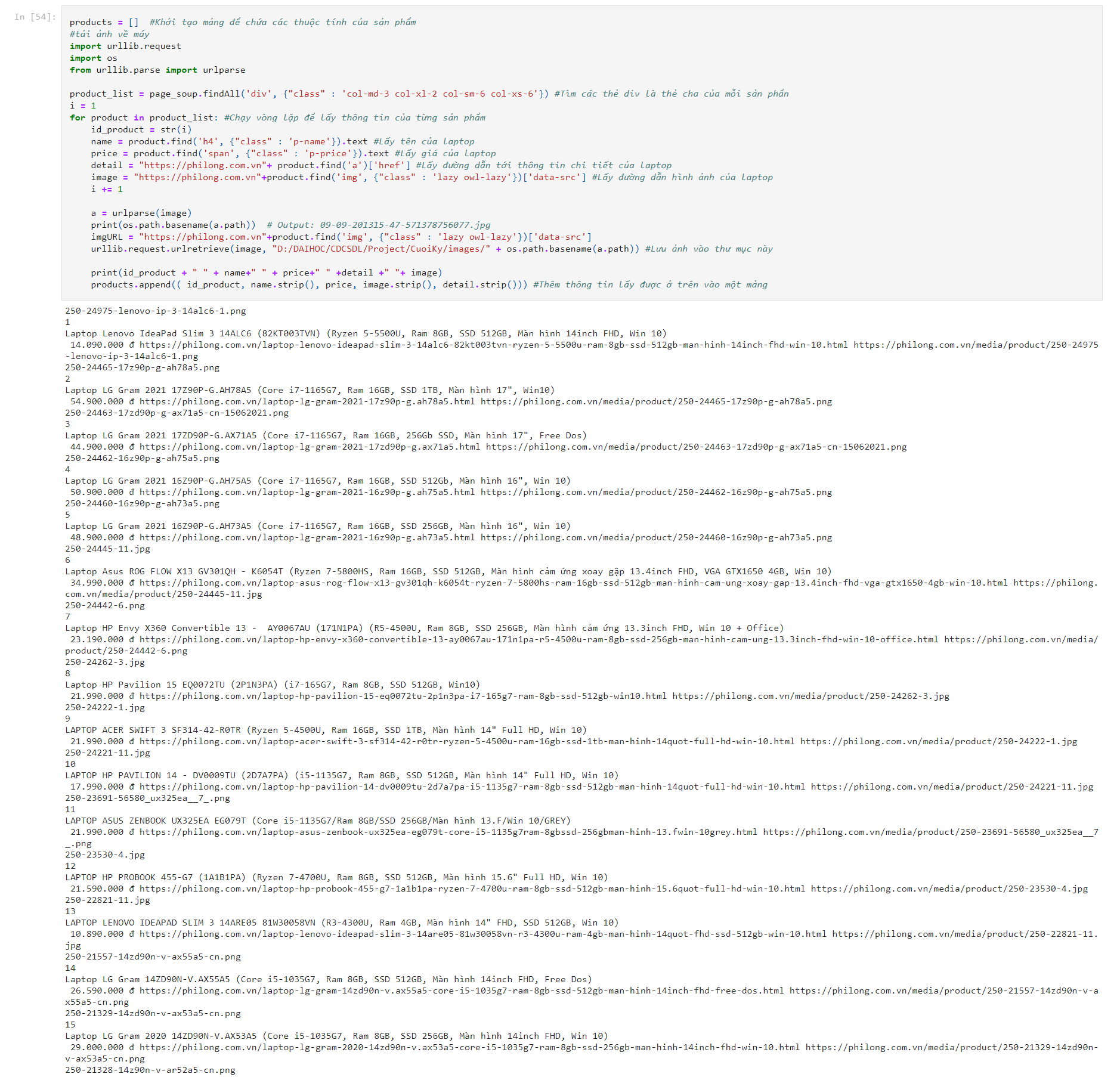




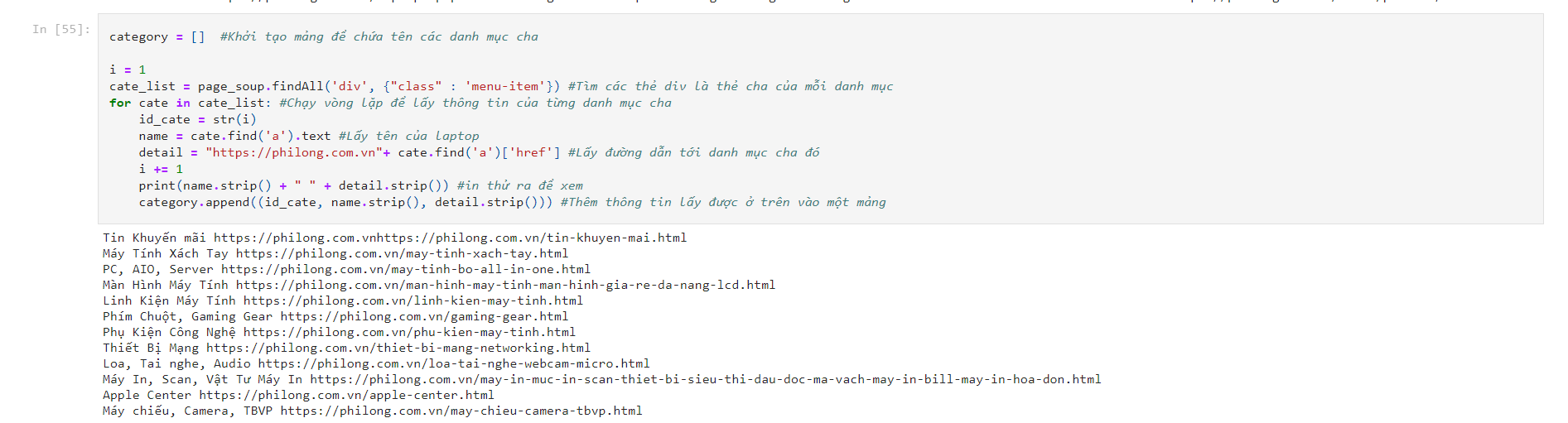
 

## Kết quả khi chạy

### Bảng sản phẩm



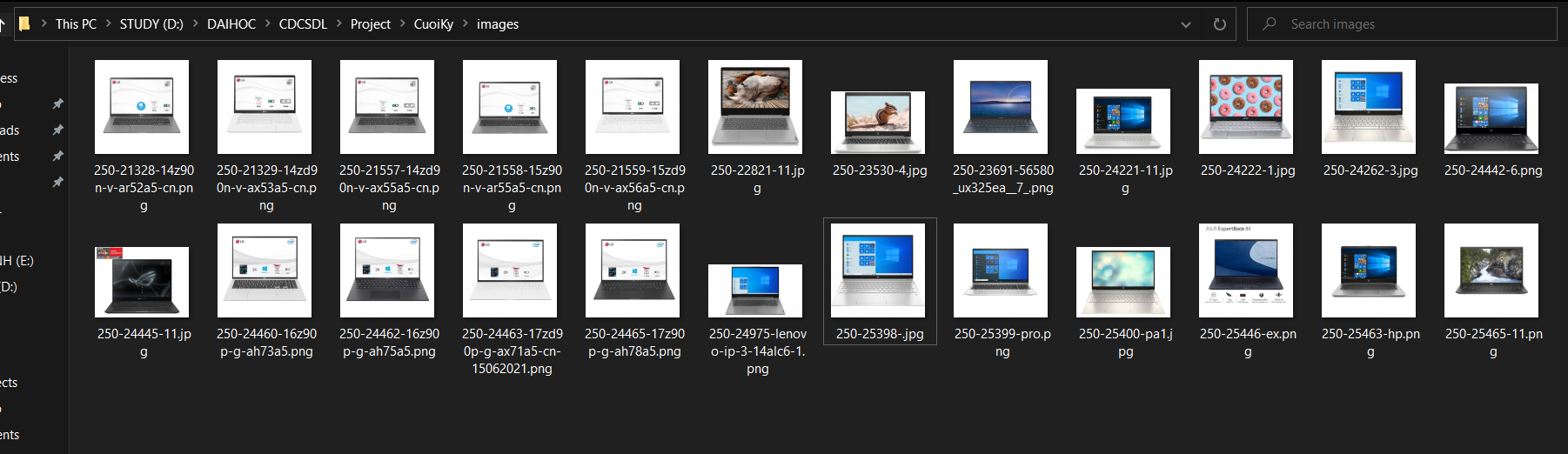
## Bảng danh mục cha



### Bảng danh mục con

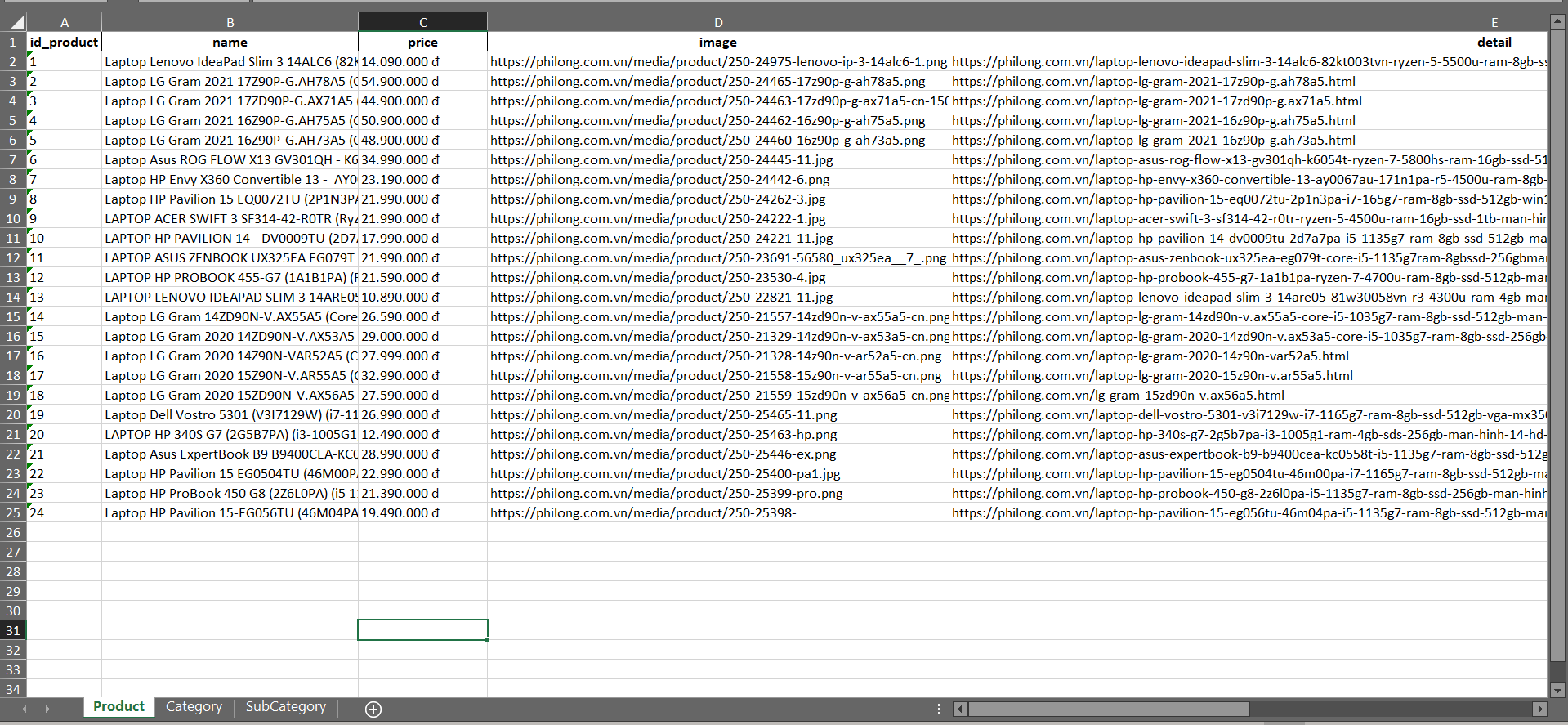


### Hình ảnh sau khi được tải về máy

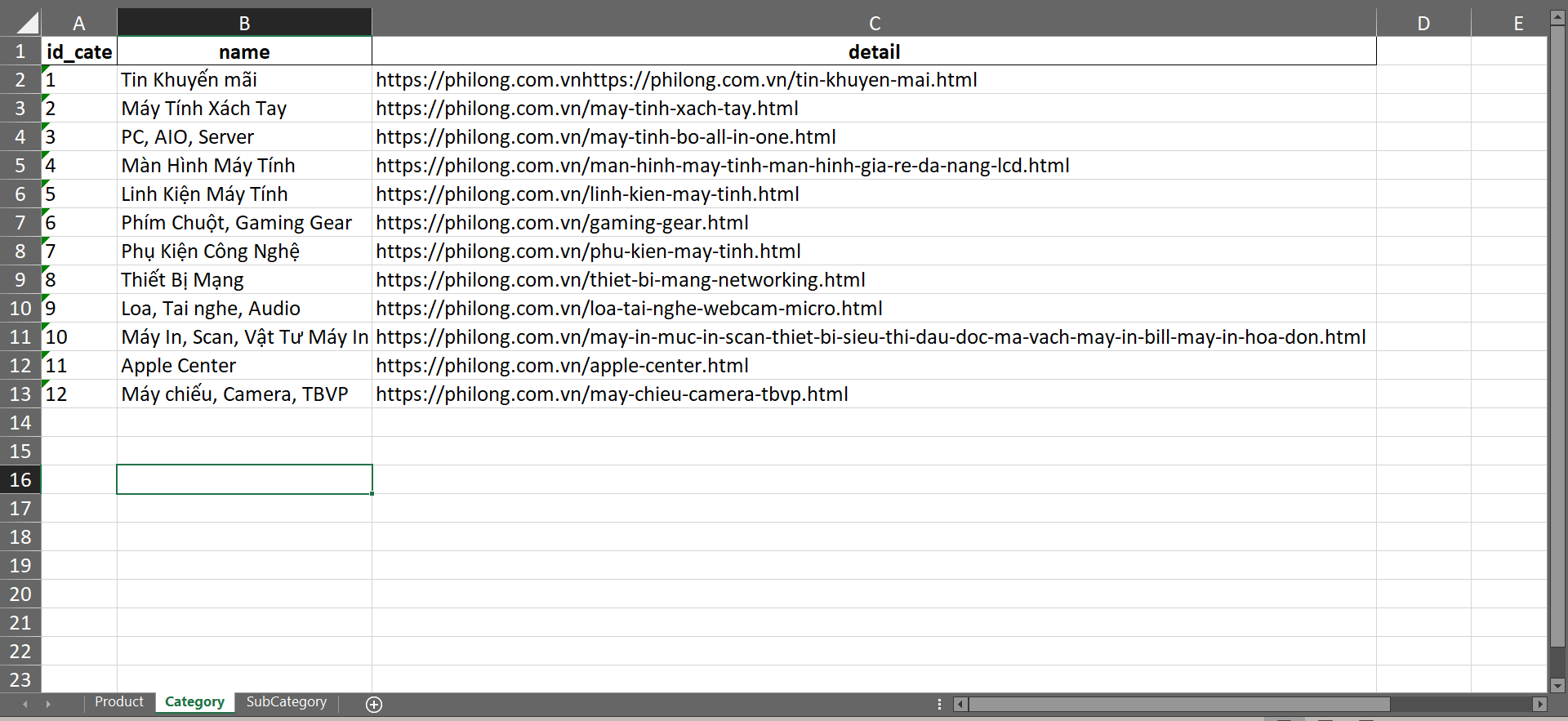


## Dữ liệu sau khi được xuất ra file exce

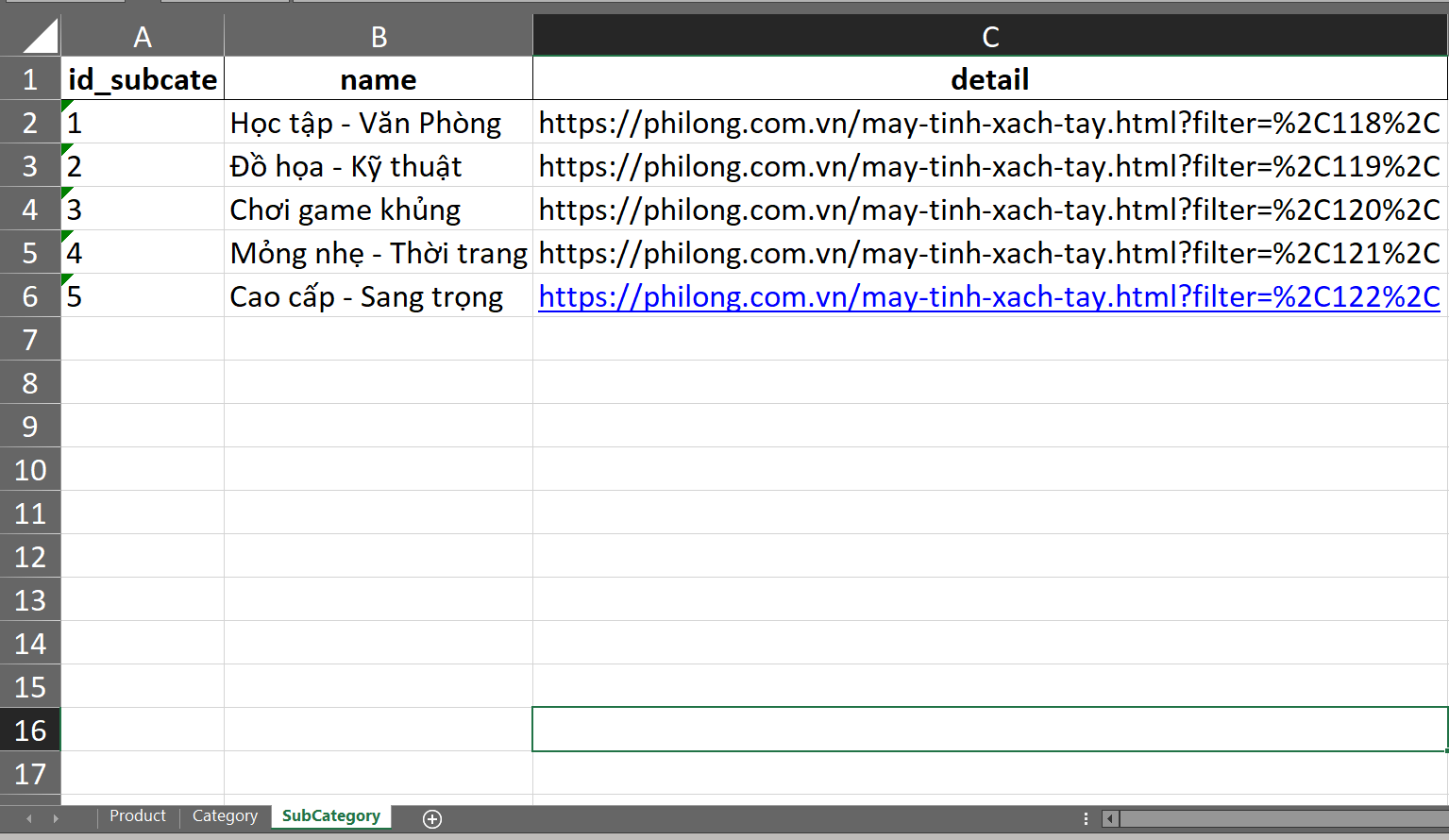
### Sheet Product



### Sheet Category



### Sheet Subcategory



## Sơ đồ ERD

