

AI Coffee - Ôn Tập SQL

Hoàng-Nguyên Vũ

Câu 1:

Sử dụng thư viện Python `requests` và `BeautifulSoup` để cào dữ liệu từ trang web chứa thông tin về diện tích, dân số, và mật độ dân số của các quận tại TP.HCM. Sau đó, lưu trữ dữ liệu này vào cơ sở dữ liệu SQLite.

Yêu cầu:

1. Cào dữ liệu từ trang web:

- Cào các thông tin bao gồm tên quận, diện tích (km²), dân số, và mật độ dân số.
- URL của trang web chứa dữ liệu: xem tại đây.

2. Lưu trữ dữ liệu vào cơ sở dữ liệu SQLite:

- Tạo cơ sở dữ liệu SQLite với bảng `Districts` có các cột:
 - `district_id`: ID tự động tăng cho từng quận.
 - `name`: Tên quận.
 - `area_km2`: Diện tích quận (đơn vị km²).
 - `population`: Dân số của quận.
 - `population_density`: Mật độ dân số (người/km²).

3. Chạy mã Python:

- Cào dữ liệu từ trang web và chèn vào bảng `Districts` trong cơ sở dữ liệu SQLite.

Hướng dẫn chi tiết:

- Sử dụng thư viện `requests` để gửi yêu cầu HTTP và nhận nội dung HTML từ trang web.
- Sử dụng thư viện `BeautifulSoup` để phân tích cú pháp HTML, tìm bảng chứa dữ liệu quận, và trích xuất các thông tin cần thiết.
- Lưu trữ dữ liệu đã cào được vào cơ sở dữ liệu SQLite thông qua thư viện `sqlite3`.

Kết quả mong đợi:

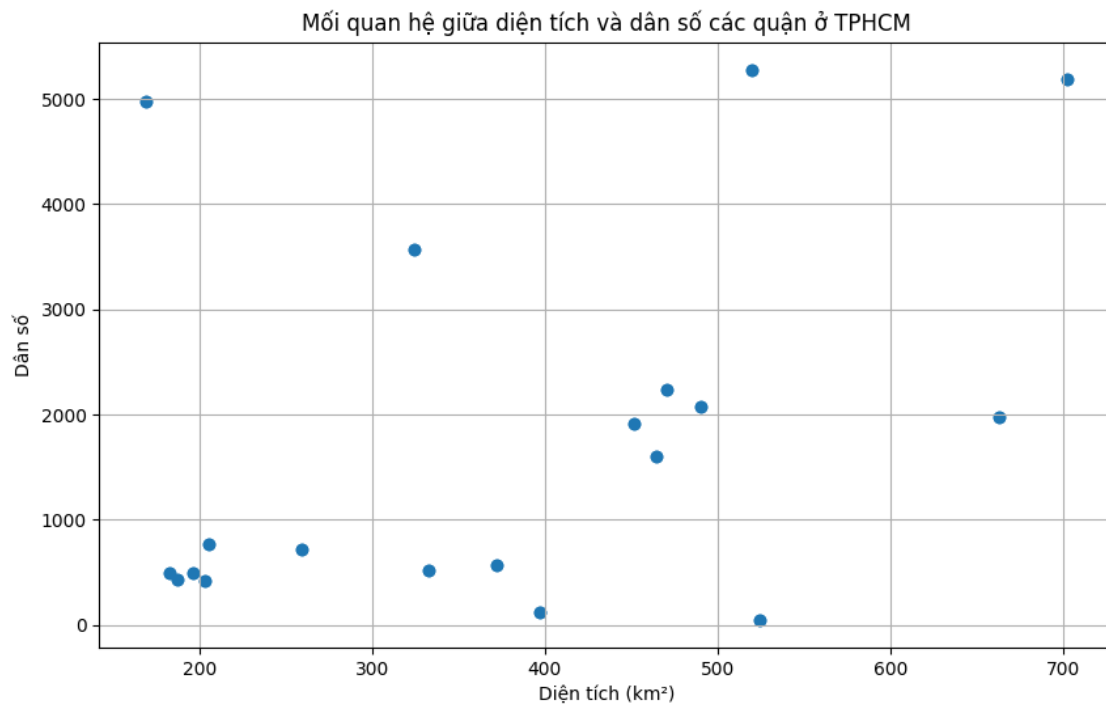
Sau khi chạy mã, bảng `Districts` trong cơ sở dữ liệu SQLite sẽ chứa thông tin về các quận của TP.HCM, bao gồm tên quận, diện tích, dân số, và mật độ dân số.

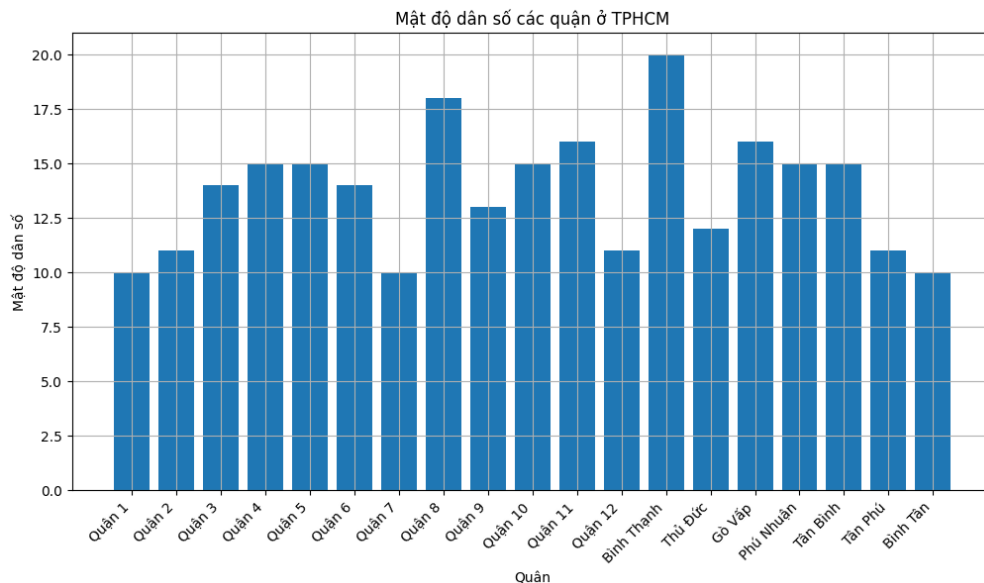
	district_id	name	area_km2	population	population_density
0	1	Quận 1	205.180	773	10
1	2	Quận 2	168.680	4974	11
2	3	Quận 3	196.433	492	14
3	4	Quận 4	203.060	418	15
4	5	Quận 5	187.510	427	15
5	6	Quận 6	258.945	719	14
6	7	Quận 7	324.620	3569	10
7	8	Quận 8	451.290	1918	18
8	9	Quận 9	397.000	114	13
9	10	Quận 10	372.450	572	15

Câu 2:

Dựa trên dữ liệu từ câu 1, các bạn hãy vẽ 2 biểu đồ để phân tích dữ liệu cơ bản về: (1) Mối quan hệ giữa diện tích và dân số của các quận tại TP.HCM và (2) Mật độ dân số của các quận tại TP.HCM.

Kết quả:





Câu 3:

1. Tạo cơ sở dữ liệu và bảng

Sử dụng SQLite để tạo các bảng sau:

- **Category**(Id, Name, Status) - Lưu thông tin danh mục sản phẩm.
- **Product**(Id, Name, Status, Price, SalePrice, CreatedDate, CategoryId) - Lưu thông tin về sản phẩm.
- **Customer**(Id, Name, Email, Phone, Address, CreatedDate, Gender, BirthDay) - Lưu thông tin khách hàng.
- **Orders**(Id, CustomerId, Status, OrderDate) - Lưu thông tin đơn hàng.
- **OrderDetail**(OrderId, ProductId, Quantity, Price) - Lưu thông tin chi tiết đơn hàng.

2. Chèn dữ liệu mẫu

Chèn các dữ liệu mẫu vào các bảng:

- **Category:** Clothes, Shoes, Accessories, Hats, Bags.
- **Product:** T-shirt, Jeans, Sneakers, Cap, Belt.
- **Customer:** John Doe, Jane Smith.
- **Orders:** Thông tin về các đơn hàng của khách hàng.
- **OrderDetail:** Thông tin chi tiết về các sản phẩm trong đơn hàng.

3. Thực hiện các truy vấn SQL và cập nhật

Viết mã Python để thực hiện các yêu cầu sau:

Yêu cầu 1: Lấy danh sách sản phẩm sắp xếp giảm dần theo Price

Truy vấn thông tin sản phẩm bao gồm Id, Name, Price, SalePrice, Status, CategoryName, CreatedDate, sắp xếp theo Price giảm dần.

Yêu cầu 2: Lấy danh sách Category và tổng số sản phẩm

Truy vấn thông tin của các Category và tính tổng số sản phẩm trong mỗi Category.

Yêu cầu 3: Truy vấn danh sách Customer và tính tuổi từ BirthDay

Truy vấn danh sách khách hàng kèm tuổi, tính toán tuổi dựa trên BirthDay.

Yêu cầu 4: Xóa các sản phẩm chưa được bán

Xóa những sản phẩm chưa từng được bán từ bảng Product.

Yêu cầu 5: Cập nhật SalePrice tăng 10% cho các sản phẩm có SalePrice \leq 250,000

Cập nhật SalePrice cho tất cả các sản phẩm có giá khuyến mãi nhỏ hơn hoặc bằng 250,000 VND, tăng thêm 10%.

Hiển thị kết quả bằng Pandas

Sử dụng thư viện Pandas để hiển thị kết quả của các truy vấn đã thực hiện:

- Hiển thị danh sách sản phẩm sắp xếp theo giá.
- Hiển thị danh sách các danh mục và tổng số sản phẩm.
- Hiển thị danh sách khách hàng và tuổi.

Kết quả:

Danh sách sản phẩm sắp xếp theo Price giảm dần:

	Id	Name	Price	SalePrice	Status	CategoryName	CreatedDate
0	3	Sneakers	1200000.0	1100000.0	1	Shoes	2024-01-03
1	2	Jeans	500000.0	450000.0	1	Clothes	2024-01-02
2	1	T-shirt	300000.0	270000.0	1	Clothes	2024-01-01
3	4	Cap	200000.0	180000.0	1	Hats	2024-01-04
4	5	Belt	150000.0	130000.0	1	Accessories	2024-01-05

Danh sách Category và tổng số sản phẩm:

	Id	Name	TotalProduct	Status
0	1	Clothes	2	1
1	2	Shoes	1	1
2	3	Accessories	1	1
3	4	Hats	1	1
4	5	Bags	0	1

Danh sách Customer và tuổi:

	Id	Name	Email	Phone	Address	CreatedDate	\
0	1	John Doe	john@example.com	123456789	123 Street	2024-01-01	
1	2	Jane Smith	jane@example.com	987654321	456 Avenue	2024-01-02	

	Gender	BirthDay	Age
0	0	1990-05-15	34
1	1	1985-08-25	39

Đã xóa các sản phẩm chưa được bán.

Đã cập nhật SalePrice cho các sản phẩm có SalePrice <= 250000.

Danh sách sản phẩm sau khi cập nhật SalePrice:

	Id	Name	Status	Price	SalePrice	CreatedDate	CategoryId
0	1	T-shirt	1	300000.0	270000.0	2024-01-01	1
1	2	Jeans	1	500000.0	450000.0	2024-01-02	1
2	3	Sneakers	1	1200000.0	1100000.0	2024-01-03	2
3	4	Cap	1	200000.0	198000.0	2024-01-04	4