

Basic Python - Data Analysis with Visualization

Hoàng-Nguyên Vũ

1. Mô tả: Làm quen với thư viện PygWalker



- **Thư viện PygWalker** là một thư viện Python mã nguồn mở giúp bạn dễ dàng chuyển đổi dữ liệu thành các ứng dụng phân tích trực quan. Thư viện này cung cấp một bộ công cụ mạnh mẽ để khám phá, tóm tắt và trực quan hóa dữ liệu của bạn, giúp bạn hiểu rõ hơn về dữ liệu và đưa ra quyết định sáng suốt hơn.
- **Điểm nổi bật của PyG Walker:**
 - + **Tạo bảng điều khiển tương tác:** PyG Walker cho phép bạn tạo các bảng điều khiển trực quan và dễ sử dụng để khám phá dữ liệu của bạn. Bạn có thể dễ dàng thêm và loại bỏ các biểu đồ, thay đổi bộ lọc và tương tác với dữ liệu theo thời gian thực.
 - + **Hỗ trợ nhiều loại biểu đồ:** PyG Walker cung cấp nhiều loại biểu đồ khác nhau để trực quan hóa dữ liệu của bạn, bao gồm biểu đồ thanh, biểu đồ đường, biểu đồ phân tán, biểu đồ nhiệt, v.v.
 - + **Khả năng lọc và nhóm dữ liệu:** PyG Walker cho phép bạn lọc dữ liệu theo các tiêu chí cụ thể và nhóm dữ liệu theo các trường khác nhau.
 - + **Tích hợp với Jupyter Notebook:** PyG Walker có thể được sử dụng trong Jupyter Notebook, cho phép bạn kết hợp phân tích dữ liệu với mã Python khác.
 - + **Dễ sử dụng:** PyG Walker có API đơn giản và dễ sử dụng, giúp bạn dễ dàng bắt đầu.
- **Ứng dụng của PygWalker trong việc trực quan hóa dữ liệu:**
 - + **Khoa học dữ liệu:** PyG Walker có thể được sử dụng để khám phá và phân tích dữ liệu trong khoa học dữ liệu.
 - + **Học máy:** PyG Walker có thể được sử dụng để chuẩn bị dữ liệu và đánh giá mô hình học máy.
 - + **Tài chính:** PyG Walker có thể được sử dụng để phân tích dữ liệu tài chính và thị trường chứng khoán.
 - + **Tiếp thị:** PyG Walker có thể được sử dụng để phân tích dữ liệu khách hàng và chiến dịch tiếp thị.

2. Cách cài đặt và sử dụng một số tính năng:

Để cài đặt thư viện PygWalker, chúng ta có thể cài trên Google Colab, hoặc ở máy cá nhân thông qua Jupyter Notebook. Cách cài đặt như sau:

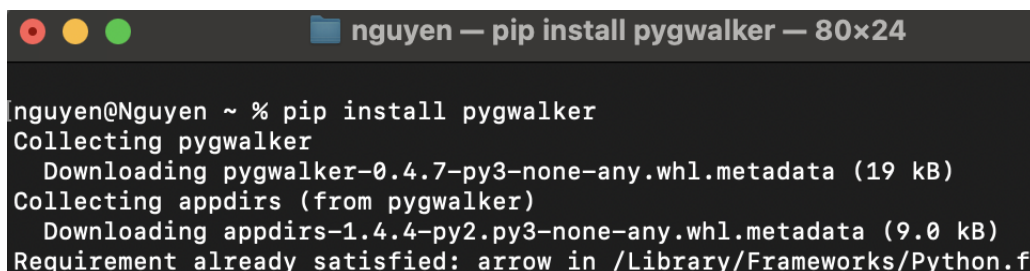
1. Cài đặt thư viện PygWalker:

+ Để cài đặt thư PygWalker, chúng ta sử dụng câu lệnh sau ở Google Colab:

```
1 !pip install pygwalker
```

+ Để cài thư viện trên máy cá nhân và chạy với Jupyter Notebook, chúng ta sẽ chạy thông qua Terminal đối với hệ điều hành MacOS và CMD đối với hệ điều hành Windows thông qua lệnh sau:

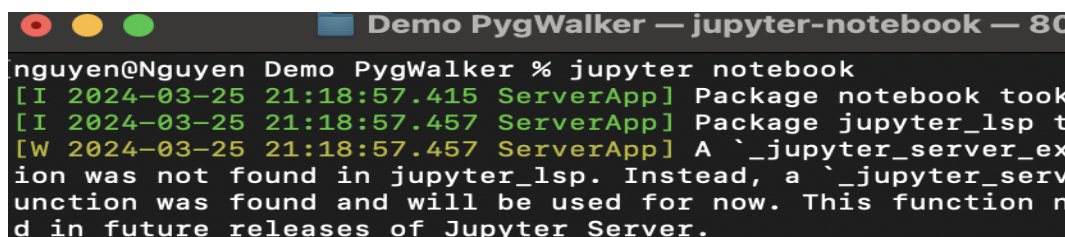
```
1 pip install pandas
2 pip install pygwalker
```



Hình 1: Cài đặt PygWalker

+ Sau khi cài xong chúng ta khởi động jupyter notebook tại thư mục chứa project của chúng ta để sử dụng, thông qua lệnh sau:

```
1 jupyter notebook
```



Hình 2: Khởi chạy Jupyter Notebook

+ Để khởi động thư viện PygWalker trong Colab/Jupyter Notebook, trước tiên ta cần phải có dataset để thư viện có thể trực quan hóa dữ liệu. Sau khi chúng ta đã có data, để khởi tạo thư viện PygWalker như sau:

```
1 import pygwalker as pyg
2 import pandas as pd
3 # FILE_PATH: đường dẫn tới tập tin CSV
4 data = pd.read_csv(FILE_PATH)
5 pyg.walk(data)
```

Showing 1 to 100 of 199 results

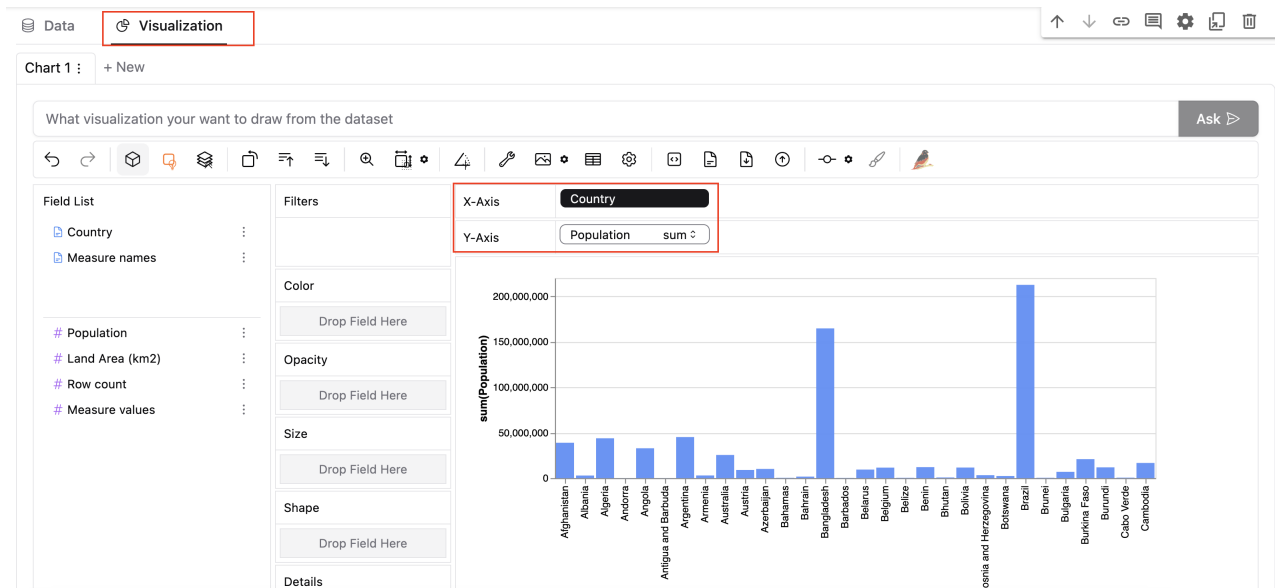
Country	Population	Land Area (km2)
199 unique values	199 unique values	700 1% 460 1% Other (195) 98%
Brazil	212,559,417	8,511,965
Brunei	437,479	5,760
Bulgaria	6,948,445	108,560
Burkina Faso	20,903,273	273,600
Burundi	11,890,784	25,680
Cabo Verde	555,987	4,030
Cambodia	16,718,965	176,520
Cameroon	26,545,863	472,710
Canada	37,742,154	9,093,510
Central African Republic	4,829,767	622,980
Chad	16,425,864	1,259,200
Chile	19,116,201	743,532
China	1,439,323,776	9,388,211
Colombia	50,882,891	1,109,500

Hình 3: Giao diện PygWalker

2. Sử dụng một số tính năng trực quan hóa:

- + **Tạo biểu đồ cột:** Tạo biểu đồ cột cho tập data mẫu trên để thể hiện dân số (Population) theo quốc gia (Country).

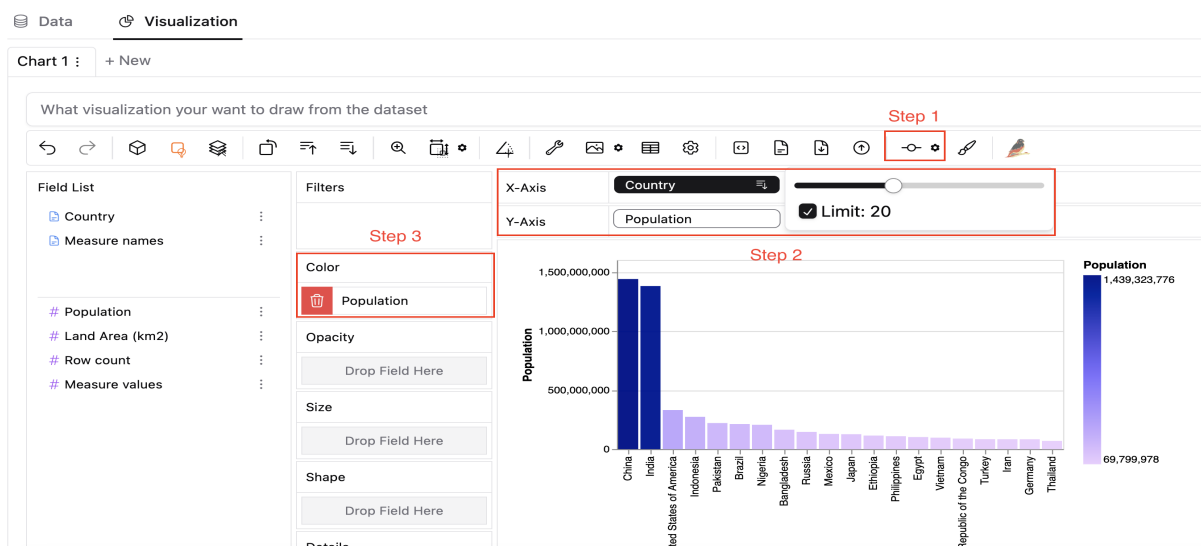
(*) Ta sẽ thực hiện kéo 2 cột: Country vào X-Axis và Population vào Y-Axis. Ứng với 2 thông số của biểu đồ cột mà chúng ta cần thực hiện trực quan hóa biểu đồ: Trục Ox (Trục ngang) thể hiện cho Quốc gia (Country) và Trục Oy (Trục dọc) thể hiện cho Dân số (Population)



Hình 4: Biểu đồ cột thể hiện dân số theo quốc gia

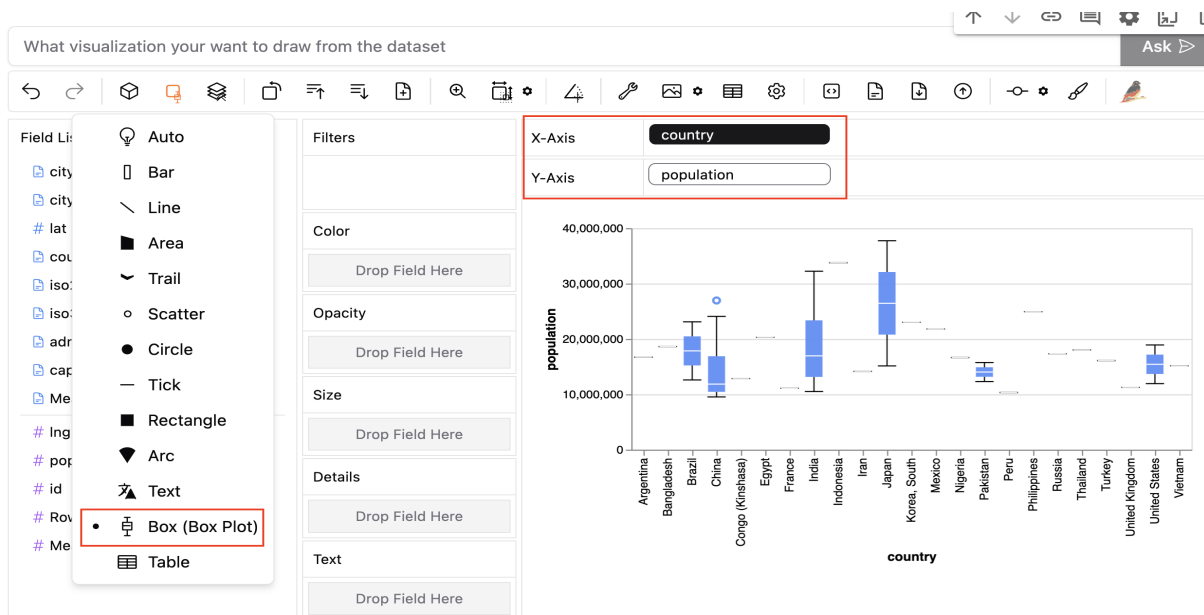
+ Lấy Top 20 Quốc Gia có giảm dần theo dân số, và tô màu theo độ lớn của dân số:

(*) Ta sẽ thực hiện kéo 2 cột: Chúng ta cũng thực hiện tương tự bài trên nhưng chúng sẽ thực hiện sắp xếp Population giảm dần và Limit 20 dòng. Đồng thời gắn Color là cột Population để thư viện hiện hiện tô màu theo Population.



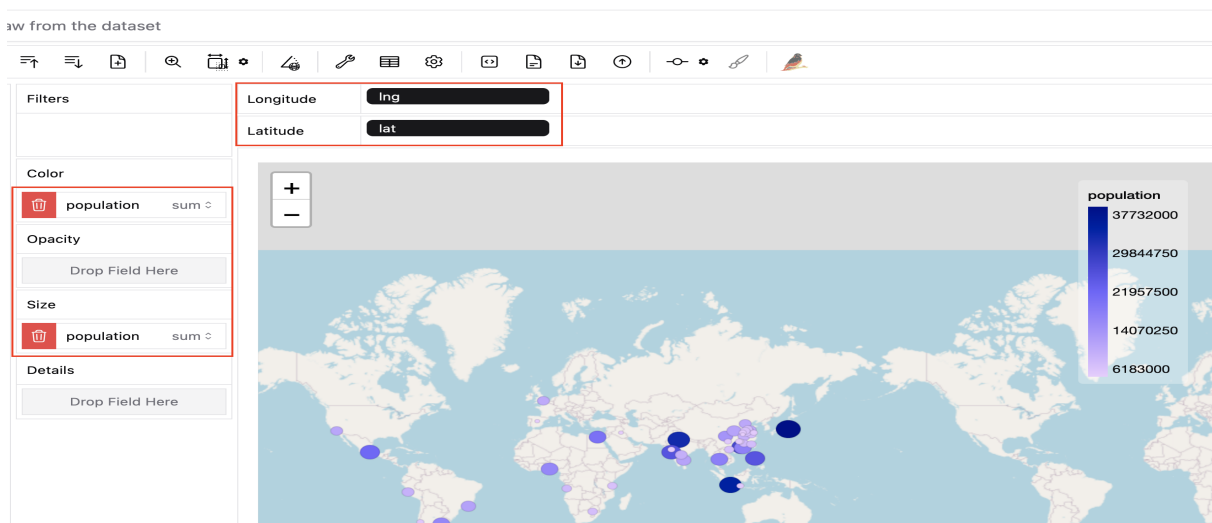
Hình 5: Biểu đồ cột thể hiện dân số theo quốc gia

+ Vẽ biểu đồ hộp thể hiện phân phối dữ liệu:



Hình 6: Biểu đồ hộp thể hiện dân số theo quốc gia

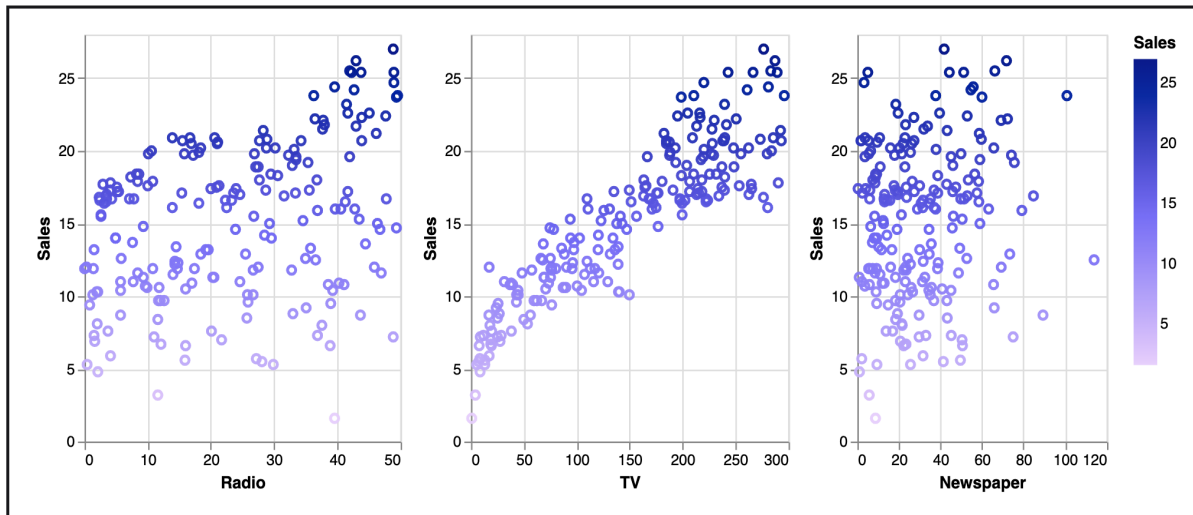
+ Vẽ bản đồ hộp thể hiện phân phối dữ liệu:



Hình 7: Bản đồ thể hiện dân số theo quốc gia

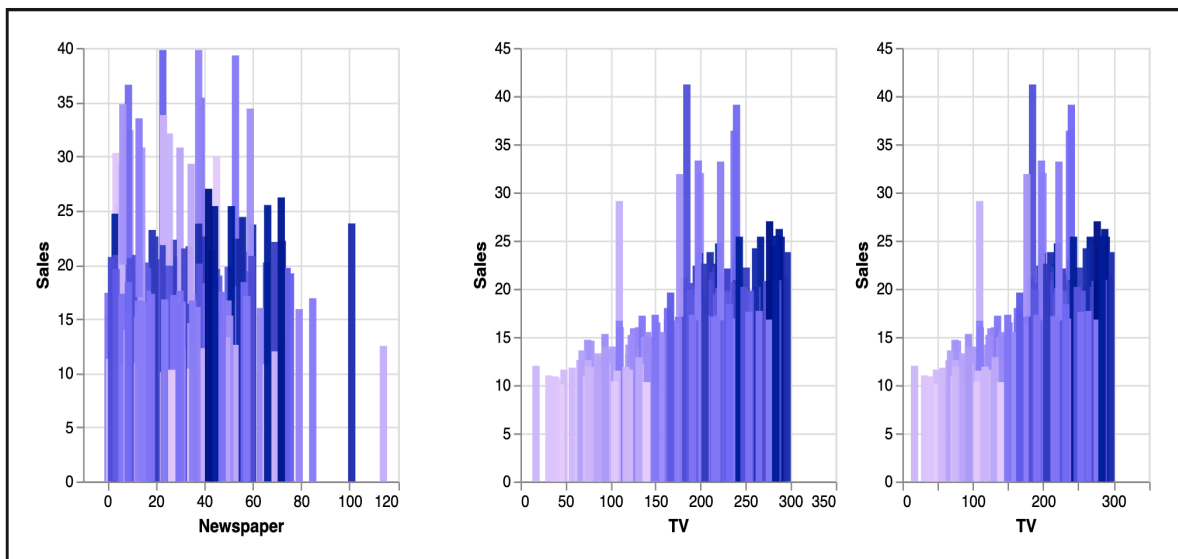
3. **Bài tập:** Hãy đọc dữ liệu ở file: advertising.csv và khởi chạy thư viện PygWalker sau đó thực hiện trực quan các biểu đồ sau đây:

- **Câu 1:** Vẽ biểu đồ phân phối dữ liệu cho 3 loại: TV, Radio, Paper và được Color theo độ lớn của giá bán (Sales).

Kết Quả:

Hình 8: Biểu đồ phân phối dữ liệu cho 3 loại: TV, Radio, Paper

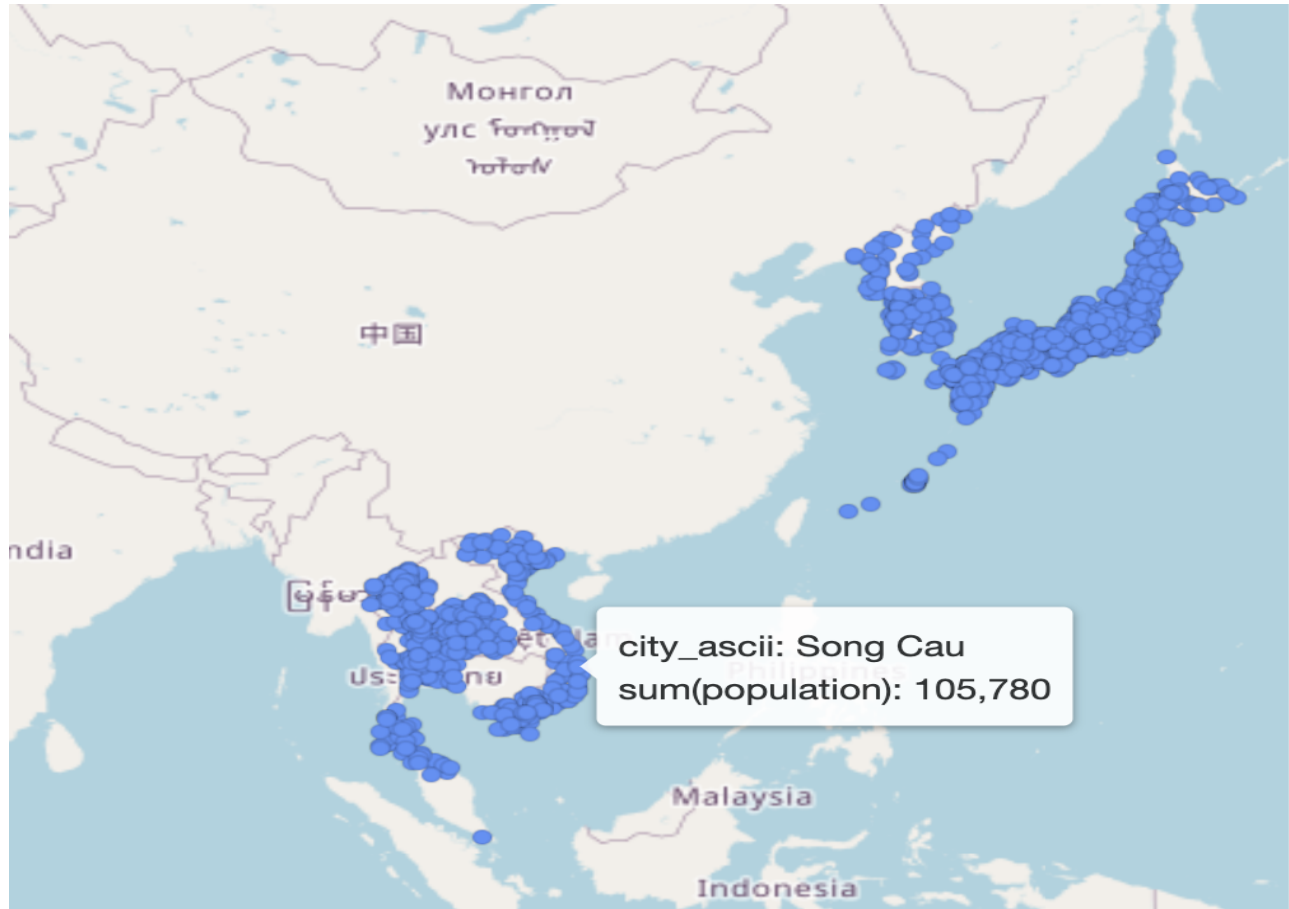
- **Câu 2:** Vẽ biểu đồ cột thể hiện doanh số bán (Sales) ≥ 10 của cả 3 loại TV, Radio và Newspaper.

Kết quả:

Hình 9: Biểu đồ doanh số bán hàng trên 10 sản phẩm cho 3 loại: TV, Radio, Paper

- **Câu 3:** Hãy vẽ bản đồ biểu diễn phân bố dân số theo thành phố của các nước thuộc các nước: Việt Nam, Hàn Quốc, Nhật Bản, Singapore và Thái Lan dựa theo dữ liệu sau: Dữ liệu dân số thế giới

Kết quả:



Hình 10: Bản đồ thể hiện phân bố dân số theo thành phố của các nước thuộc các nước: Việt Nam, Hàn Quốc, Nhật Bản, Singapore và Thái Lan

- Hết -