### **Giới thiệu thư viện seaborn**

### **Chi tiết bài đọc**

**Giới thiệu**

Matplotlib là một thư viện dùng để trực quan hóa rất phổ biến cho Python. Tuy nhiên, matplotlib còn có rất nhiều thiếu sót. Vấn đề đầu tiên, API của Matplotlib là low-level, chúng ta có thể thực hiện trực quan hóa thống kê phức tạp, nhưng thường đòi hỏi rất nhiều tham số, xử lý, dẫn đến phải dùng các đoạn code mẫu. Vấn đề tiếp theo, Pandas xuất hiện sau Matplotlib rất lâu, do đó không được thiết kế để sử dụng với Pandas DataFrames. Để visualizing một pandas dataframe, ta phải trích xuất từng dòng và nối chúng lại với nhau thành đúng định dạng. Sẽ thuận tiện hơn nếu có một thư viện thông minh sử dụng trực tiếp DataFrame trong khi vẽ. Mặc dù những phiên bản gần đây, Matplotlib đang cố gắng giải quyết những vấn đề này, tuy nhiên vẫn còn nhiều thiếu sót.

Seaborn ra đời để giải quyết những vấn đề trên. Seaborn cung cấp các APIs xây dựng trên thư viện Matplotlib, cung cấp các lựa chọn cho các kiểu mặc định và màu sắc, định nghĩa các chức năng cấp cao đơn giản cho các kiểu thống kê chung và tích hợp với các chức năng do Pandas DataFrames cung cấp.

Một số tính năng của Seaborn:

* Một số các themes được dựng sẵn để tạo styles cho hình vẽ trong matplotlib.
* Công cụ để chọn bảng màu để tạo ra các đồ thị đẹp thể hiện các mẫu trong dữ liệu của bạn.
* Các hàm để hiển thị các phân bố đơn và hai biến hoặc để so sánh chúng giữa các tập con của dữ liệu.
* Các công cụ phù hợp và biểu diễn các model hồi quy tuyến tính (linear regression) cho các loại biến độc lập và phụ thuộc khác nhau.
* Các hàm biểu diễn các ma trận dữ liệu và sử dụng các thuật toán phân cụm (clustering) để khám phá cấu trúc trong những ma trận đó.
* Một chức năng để đánh số liệu thống kê timeseries với ước tính linh hoạt và đại diện cho sự không chắc chắn xung quanh ước tính.
* Cho phép ta dễ dàng xây dựng các hình ảnh phức tạp qua High-level abstractions cho cấu trúc grid của các plots.

Seaborn nên được coi là một tool bổ sung cho matplotlib chứ không phải là một sự thay thế cho nó. Khi sử dụng seaborn, rất có thể chúng ta sẽ thường gọi các hàm matplotlib trực tiếp để vẽ các đồ thị đơn giản đã có sẵn qua không gian tên pyplot. Hơn nữa, các chức năng trên Seaborn nhằm mục đích tạo ra các plots có mẫu sẵn, hợp lý (bao gồm khai thác thông tin ngữ nghĩa từ các đối tượng Pandas để thêm các nhãn thông tin), nhưng tuỳ biến đầy đủ sẽ yêu cầu thay đổi trực tiếp các thuộc tính trên các đối tượng matplotlib. Sự kết hợp giữa sự tiện dụng seaborn và khả năng tuỳ biến của matplotlib và các phụ trợ dễ dàng tạo ra các figures chất lượng cao.

**Cài đặt**

Việc cài đặt seaborn cũng giống như các cài đặt các thư viện khác trong python:

>> pip install seaborn

**Ví dụ:**

>> import matplotlib.pyplot as plt

>> import seaborn as sns #khai báo thư viện

>> sns.set\_theme(style="darkgrid")#chọn theme cho biểu đồ

>> x = [1, 2, 3, 4, 5]

>> y = [1, 5, 4, 7, 4]

>> sns.lineplot(x, y) #vẽ biểu đồ đường theo x, y

>> plt.show() #hiển thị biểu đồ

Kết quả:

Chart, line chart

Description automatically generated

Seaborn cung cấp sẵn cho ta rất nhiều bộ dữ liệu mà ta có thể load Dataframe trực tiếp từ seaborn thông qua hàm load\_dataset().

>> tips = sns.load\_dataset("tips")

>> type(tips)

pandas.core.frame.DataFrame

>> tips.head()

Table

Description automatically generated

### **Tổng kết**

Qua bài đọc này, chúng ta đã tìm hiểu về lí do ra đời của Seaborn, công dụng, cách cài đặt và ví dụ đơn giản với nó.