

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT HƯNG YÊN
KHOA CÔNG NGHỆ THÔNG TIN

----- oOo -----



BÀI TẬP THỰC HÀNH

LẬP TRÌNH PYTHON NÂNG CAO

TRÌNH ĐỘ: ĐẠI HỌC CHÍNH QUY

NGÀNH ĐÀO TẠO: KHOA HỌC MÁY TÍNH

Hưng Yên – Tháng 11 năm 2021

BÀI THỰC HÀNH SỐ 5: TRỰC QUAN HÓA DỮ LIỆU VỚI MATPLOTLIB (2)

A. MỤC TIÊU BÀI THỰC HÀNH

Sau bài thực hành này sinh viên có thể:

- Biết cách sử dụng thư viện Matplotlib để trực quan hóa dữ liệu trong Python
- Sử dụng thành thạo các hàm thao tác với dữ liệu mảng, DataFrame trong Pandas
- Tích hợp vẽ biểu đồ Line, Bar tạo thành các Subplot
- Sử dụng các hàm cơ bản của Seaborn để trực quan hóa dữ liệu
- Vận dụng các kỹ thuật xử lý dữ liệu với Pandas và Matplotlib để phân tích và trực quan hóa dữ liệu trong các bài toán thực tế

B. ĐIỀU KIỆN THỰC HÀNH

Với đặc thù của môn Lập trình Python nâng cao, mục này sẽ liệt kê một số công cụ sử dụng để làm bài thực hành. Trong bài thực hành này, sinh viên cần kiểm tra và chắc chắn các phần mềm sau trên máy tính còn hoạt động tốt:

1. Anacoda phiên bản 3.0 trở lên
2. Jupyter Notebook

C. TÀI NGUYÊN THAM CHIẾU

Để hoàn thành tốt bài thực hành này sinh viên nên tham khảo các tài nguyên sau:

STT	Tên tài nguyên	Mô tả tài nguyên
1	Practice 05.pdf	Tài liệu bài thực hành số 5
2	Lesson 05 + 06 – Data Visualization with Matplotlib.pdf	Slide bài giảng về thao Trực quan hóa dữ liệu với Matplotlib

D. YÊU CẦU BÀI THỰC HÀNH

Bài 1. Cho dữ liệu của 82,496 căn nhà của Thành phố Hà Nội được lưu trong tệp Hanoi_Housing.csv. Bạn hãy sử dụng Pandas và Matplotlib thực hiện các yêu cầu sau:

- a) Đọc dữ liệu từ tệp ra DataFrame và hiển thị ra màn hình.

- b) Chuẩn hóa dữ liệu để tạo ra DataFrame mới sau cho:
 - Bỏ đơn vị m^2 ở cột “Diện tích” và chuyển thành số
 - Bỏ đơn vị **triệu/m²** ở cột “Giá/m²” và chuyển thành số
- c) Vẽ biểu đồ Bar theo chiều ngang thể hiện top 10 Quận/Huyện có nhiều nhà nhất.
- d) Vẽ biểu đồ Pie thể hiện tỉ lệ tình trạng về “Giấy tờ pháp lý” của các căn nhà:
 - Đã có sổ
 - Đang chờ sổ
 - Giấy tờ khác
 - NaN
- e) Vẽ biểu đồ Pie thể hiện tỉ lệ nhà tầng của các căn nhà Hà Nội, gồm các mức:
 - [1, 10)
 - [10, 20)
 - [20, 30)
 - [30, 40)
 - [40, 50)
 - [50, ∞)
 - NaN
- f) Vẽ biểu đồ Histogram thể hiện sự phân bố về diện tích các căn nhà tại Hà Nội
- g) Vẽ biểu đồ Histogram thể hiện sự phân bố số lượng phòng ngủ của các căn nhà tại Hà Nội.
- h) Vẽ biểu đồ Histogram thể hiện sự phân bố về giá/m² của các căn nhà tại Hà Nội
- i) Vẽ biểu đồ thống kê thông tin các căn nhà tại Hà Nội vào một hình bằng cách sử dụng **Subplot** với 2 hàng 3 cột đưa các biểu đồ từ yêu cầu c) đến h) làm các axes.

Bài 2. Dữ liệu 2490 sản phẩm của Lazada Vietnam được lưu trong tệp LazadaProducts.json.

Bạn hãy sử dụng **Seaborn** thực hiện các yêu cầu sau:

- a) Vẽ biểu đồ Bar theo chiều dọc thể hiện số lượng sản phẩm mỗi loại.
- b) Vẽ biểu đồ Pie thể hiện tỉ lệ số lượng sản phẩm mỗi loại.
- c) Vẽ biểu đồ thống kê thông tin số lượng sản phẩm mỗi loại vào 1 hình gồm 2 hình nhỏ đã vẽ ở yêu cầu a) và b) theo chiều ngang

Bài 3: Dữ liệu chứng khoán 5 năm của các doanh nghiệp được lưu trong tệp `Stocks_5years.csv` (2/2013 - 2/2018) bao gồm các trường:

- Date: ngày
- Open: giá mở phiên
- High: Giá cao nhất trong ngày
- Low: Giá thấp nhất trong ngày
- Close: Giá đóng phiên
- Volume: Lượng cổ phiếu giao dịch trong ngày
- Name: Mã cổ phiếu

Bạn hãy lập trình thực hiện các yêu cầu sau:

- a) Vẽ biểu đồ Line với Subplot (2x2) thể hiện về “Giá đóng phiên” các ngày trong 2017 của 4 mã cổ phiếu: GOOGL (Google), AAPL (Apple), MSFT (Microsoft), AMZN (Amazon)
- b) Vẽ biểu đồ Line với Subplot (2x2) thể hiện “Giá đóng phiên” và các “Đường trung bình” (Moving Average - MA) về giá gồm: MA10, MA20, MA50 trong năm 2017 của 4 mã cổ phiếu: GOOGL (Google), AAPL (Apple), MSFT (Microsoft), AMZN (Amazon). Các đường giá của một mã cổ phiếu được vẽ trên một subplot.

Biết:

MA10 là đường hiển thị giá trị trung bình của 10 phiên (ngày) trước đó.

MA20 là đường hiển thị giá trị trung bình của 20 phiên (ngày) trước đó.

MA50 là đường hiển thị giá trị trung bình của 50 phiên (ngày) trước đó.

E. HƯỚNG DẪN THỰC HIỆN

Sinh viên tạo tệp **Practice05_HoVaTen.ipynb** trên Jupyter Notebook và thực hiện viết mã lệnh để giải quyết các bài tập thực hành. Các biểu đồ cần thể hiện các Title, Axis, Labels và Legends.

Bài 1:

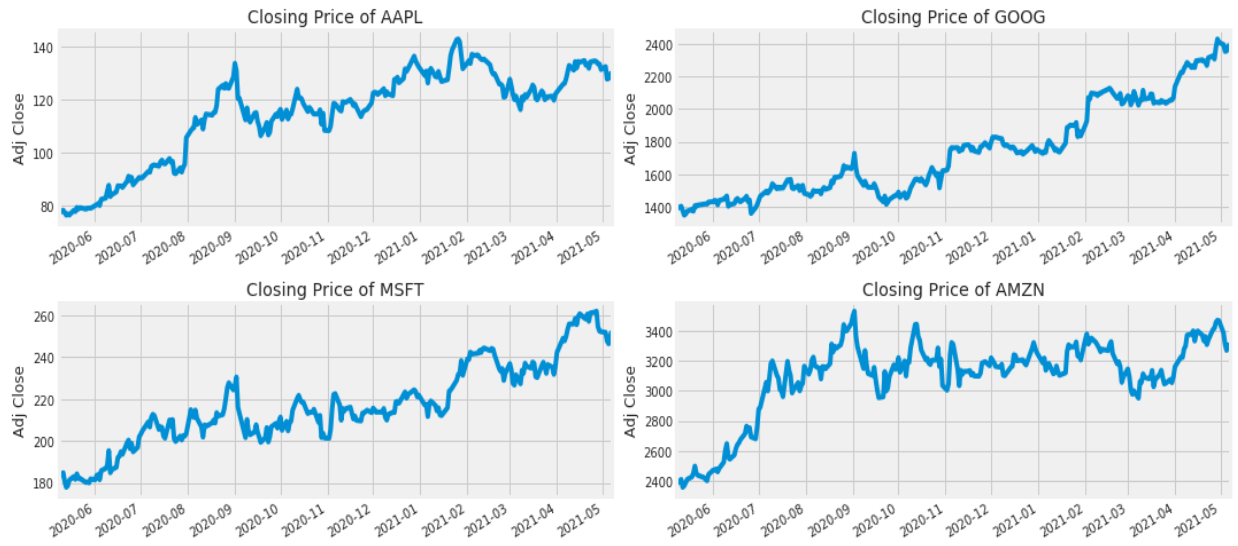
Sinh viên giải các bài tập theo ví dụ mẫu trên lớp.

Bài 2:

Sinh viên giải các bài tập theo ví dụ mẫu trên lớp.

Bài 3:

a) Hình minh họa (không phải dữ liệu trong bài)



b) Hình minh họa (không phải dữ liệu trong bài)

