

**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 4: TRỰC QUAN HÓA DỮ LIỆU VỚI MATPLOTLIB**

### **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
- Thành thạo vẽ biểu đồ Line, Bar, Histogram, Pie, Scatter
- 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 04.pdf	Tài liệu bài thực hành số 4
2	Lesson 05 – Data Visualization with Matplotlib (1).pdf	Slide bài giảng về thao Trực quan hóa dữ liệu với Matplotlib

### **D. YÊU CẦU BÀI THỰC HÀNH**

**Bài 1.** Dữ liệu 288 bệnh nhân nhiễm Covid-19 ở Việt Nam Pandas (05-2020) được lưu trong tệp Vietnam\_Covid19\_Patients.csv. Bạn hãy sử dụng Matplotlib thực hiện các yêu cầu sau:

- a) Vẽ biểu đồ Pie biểu diễn tỉ lệ bệnh nhân Nam, Nữ trong dữ liệu.
- b) Vẽ biểu đồ Bar thể hiện số lượng bệnh nhân các tháng (Tháng 1 đến Tháng 5).
- c) Vẽ biểu đồ Bar theo chiều ngang thể hiện Top 10 bệnh viện có bệnh nhân nhiều nhất.

- d) Vẽ biểu đồ Bar theo chiều ngang thể hiện Top 10 tỉnh thành có bệnh nhân nhiều nhất.
- e) Vẽ biểu đồ Line thể hiện số lượng bệnh nhân các tháng (Tháng 1 đến Tháng 5).
- f) Vẽ biểu đồ Scatter thể hiện sự phân bố số lượng bệnh nhân theo độ tuổi.
- g) Vẽ biểu đồ Histogram thể hiện sự phân bố số lượng bệnh nhân theo độ tuổi.

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

Bạn hãy sử dụng Matplotlib 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 Bar theo chiều ngang cho top 10 sản phẩm có điểm đánh giá trung bình cao nhất.
- c) Vẽ biểu đồ Pie thể hiện tỉ lệ số lượng sản phẩm mỗi loại.

**Bài 3:** Dữ liệu 8807 video clips của NetFlix được lưu trữ trong tệp “netflix\_data.xml”.

Bạn hãy lập trình thực hiện các yêu cầu sau:

- a) Vẽ biểu đồ Line biểu diễn sự tăng trưởng số Video Clip qua các năm từ 2000 đến 2022.
- b) Vẽ biểu đồ Bar biểu diễn số Video Clip được sản xuất của từng năm từ 2017 đến 2021.
- c) Vẽ biểu đồ BarGroup biểu diễn số TV Shows và Movie được sản xuất từ năm 2017 đến 2021.
- d) Vẽ biểu đồ Histogram biểu diễn sự phân bố số lượng của các chương trình qua từng năm.
- e) Vẽ biểu đồ Pie biểu diễn tỉ lệ phần trăm TV Show với tỉ lệ Movie.

## E. HƯỚNG DẪN THỰC HIỆN

Sinh viên tạo tệp **Practice04\_HoVaTen.ipynb** trên Jupyter Notebook và thực hiện viết mã lệnh để giải quyết các bài tập thực hành.

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

Các biểu đồ cần thể hiện các Axis, Labels và Legends.