



HCMUTE

TRƯỜNG ĐẠI HỌC

SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH

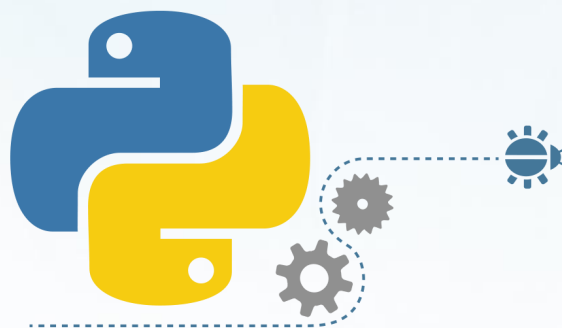
HCMC University of Technology and Education



KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN HỆ THỐNG THÔNG TIN

NHẬP MÔN LẬP TRÌNH PYTHON (IPPA233277)

BÀI TẬP THƯ VIỆN PANDAS



GV. Trần Quang Khải

- Giới thiệu thư viện pandas
- Thao tác với Series
- Thao tác với DataFrame



1. Sử dụng bộ dữ liệu Churn Customer, thực hiện các yêu cầu sau”
 - a. Đọc dữ liệu từ tập tin đã cho, có phần header là row thứ nhất và index là trường RowNumber.
 - b. Thống kê mô tả đối với các trường trong bảng dữ liệu này.
 - c. Tính trung bình điểm CreditScore theo Geography.
 - d. Phân đều Age thành 05 nhóm độ tuổi sao cho mỗi nhóm chiếm 20% số quan sát.
 - e. Vẽ biểu đồ barchart thống kê số lượng khách hàng theo nhóm độ tuổi vừa tạo được.



2. Sử dụng bộ dữ liệu **chipotle.tsv** thực hiện các yêu cầu sau:
- a. Đọc dữ liệu từ tập tin đã cho
 - b. Liệt kê những sản phẩm có giá hơn 10\$ (lược bỏ những dòng trùng tên sản phẩm)
 - c. Sắp xếp các sản phẩm theo tên
 - d. Tìm sản phẩm có giá cao nhất trong danh sách
 - e. Cho biết sản phẩm “Veggie Salad Bowl” xuất hiện trong bao nhiêu đơn hàng với tổng số lượng được đặt
 - f. Vẽ biểu đồ histogram cho 5 sản phẩm được mua nhiều nhất với tần suất mua
 - g. Vẽ biểu đồ scatter với số lượng mặt hàng được đặt hàng trên mỗi đơn hàng



3. Sử dụng bộ dữ liệu **u.user** thực hiện các yêu cầu sau:
- a. Đọc dữ liệu từ tập tin đã cho
 - b. Cho biết độ tuổi trung bình của mỗi nghề nghiệp
 - c. Cho biết tỷ lệ nam trên mỗi nghề và sắp xếp từ cao đến thấp
 - d. Với mỗi nghề nghiệp, hãy cho biết độ tuổi nhỏ nhất và lớn nhất
 - e. Với mỗi tổ hợp của nghề nghiệp – giới tính, hãy tính tuổi trung bình
 - f. Với mỗi nghề nghiệp, hãy cho biết tỷ lệ phần trăm của nam – nữ



4. Thực hiện các yêu cầu sau:

- a. Tạo ba Series khác nhau có kích thước 100
 - ✓ Giá trị ngẫu nhiên trong đoạn [1 .. 4]
 - ✓ Giá trị ngẫu nhiên trong đoạn [1 .. 3]
 - ✓ Giá trị ngẫu nhiên trong đoạn [10000 .. 30000]
- b. Tạo một DataFrame bằng nối các Series theo cột
- c. Đổi tên các cột thành `bedrs`, `bathrs`, `price_sqr_meter`
- d. Tạo một cột DataFrame với giá trị của các Series, gán 'bigcolumn'
- e. Lập chỉ mục lại cho DataFrame ở trên chuyển từ 0 → 299



5. Sử dụng bộ dữ liệu **wind.data** thực hiện các yêu cầu sau:
 - a. Đọc dữ liệu từ tập tin đã cho
 - b. Hợp nhất ba cột đầu tiên (Yr, Mo, Dy) bằng giá trị định dạng “yyyy-MM-dd” với tên cột “Yr_Mo_Dy”
 - c. Đặt cột mới vừa phát sinh ở câu trên làm chỉ mục
 - d. Hãy cho biết số lượng giá trị hiện có và còn thiếu ở các cột từ RPT → MAL
 - e. Tính tốc độ giá trung bình của toàn bộ dữ liệu (mọi nơi và mọi thời điểm)
 - f. Tạo một DataFrame có tên là loc_stats và tính tốc độ gió tối thiểu, tối đa, trung bình, độ lệch chuẩn của tốc độ giá tại mỗi vị trí trong tất cả các ngày thống kê
 - g. Tìm tốc độ gió trung bình trong tháng 1 ở mỗi nơi
 - h. Giảm dữ liệu, thống kê theo từng năm, theo tháng – năm, theo tuần – tháng – năm



- ✓ Họ tên : **Trần Quang Khải**
- ✓ Email : **khaitq@hcmute.edu.vn**
- ✓ Zalo (mã Qr)

