

## Bài tập 6

**Bài 1.** Hãy tạo 2 series mỗi series có 5 phần tử là các số nguyên ngẫu nhiên. Thực hiện các phép toán giữa 2 series. Vẽ biểu đồ tương quan giữa 2 series này.

**Bài 2.** Hãy tạo một DataFrame có 5 cột và 6 dòng là các số nguyên nhỏ hơn 100. Hãy chuẩn hóa dữ liệu trong DataFrame về miền  $[0,1]$ . Tính giá trị lớn nhất, nhỏ nhất của các cột. Vẽ biểu đồ so sánh giá trị trung bình giữa các cột.

**Bài 3.** Cho bộ dữ liệu về Tỷ giá ngoại tệ mua vào so với đồng VNĐ như sau:

Mã NT	AUD	CAD	CHF	EUR	GBP	HKD	JPY	SGD	THB	USD
Mua tiền mặt	16.815	17.138	22.568	26.172	29.876	2.864	201	16.734	692	22.775

- Lưu bộ dữ liệu trên ở file `tigia.csv`, có 3 cột: Mã NT, Mua tiền mặt
- Đọc file `tigia.csv` đã tạo ở câu 1, in nội dung dữ liệu đọc được ra màn hình
- Sắp xếp Mã NT theo thứ tự Alphabet và lưu vào file `tigia2.csv`
- Vẽ biểu đồ cột chiều đứng thể hiện tương quan về tỷ giá của các ngoại tệ với VNĐ, yêu cầu thêm chú thích phù hợp cho các dòng và cột trên biểu đồ để có thể truyền đạt nguyên vẹn các thông tin đến người sử dụng biểu đồ.
- Lưu biểu đồ đã vẽ ở Câu d ra file `tigia.png`

## Bài 4 Hãy thực hiện các công việc sau:

a) Hãy tạo một bộ dữ liệu có file “K59.csv” có các cột sau: `soluongmau`, `KNN`, `FIS`. Có nội dung như sau:

<code>soluongmau</code>	<code>KNN</code>	<code>FIS</code>
300	0.904	0.912
500	0.893	0.901
700	0.863	0.863
1000	0.823	0.812
1200	0.802	0.794
1500	0.783	0.743
1800	0.765	0.732
2000	0.752	0.723

- Hãy đọc file “K60.csv” đã tạo ở trên, in nội dung ra màn hình
- Hãy vẽ biểu đồ đường so sánh sự tương quan giữa 2 phương pháp KNN, FIS.
- Xuất biểu đồ đã vẽ ở trên ra file “K59.png”
- Hãy tìm hệ số tương quan hồi quy và sai số giữa phương pháp KNN với phương pháp FIS.

**Bài 5** Tải về file `winequality.csv` về các số đo của rượu vang và chất lượng của rượu  
Liên kết: <http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-red.csv>

- Đây là bộ data của đại học California-Berkeley
- Bộ data gồm 1599 mẫu rượu vang, mỗi mẫu gồm 11 loại chỉ số và đánh giá của chuyên gia về chất lượng rượu (cột `quality`, điểm số từ 0 đến 10)

Chú ý:

- Dữ liệu sử dụng dấu chấm phẩy (;) để ngăn giữa các cột

- Tên các cột có chứa dấu cách

**In ra dữ liệu vừa tải về, ý nghĩa các cột thuộc tính**

▪ fixed acidity	Nồng độ axit tartaric
▪ volatile acidity	Tính axit
▪ citric acid	Nồng độ axit Citric
▪ residual sugar	Nồng độ đường dư
▪ chlorides	Nồng độ clo
▪ free sulfur dioxide	Nồng độ acid sulfuric tự do
▪ total sulfur dioxide	Nồng độ acid sulfuric
▪ density	Mật độ (khối lượng/đơn vị thể tích)
▪ pH	Độ pH
▪ sulphates	Nồng độ sunfat
▪ alcohol	Nồng độ chất alcohol

a) Hãy đọc file “winequality-red.csv”.

b) Tính tổng, min, max, trung bình của các cột.

c) Sử dụng các cột thuộc tính “alcohol” để tìm tương quan giữa thuộc tính này và điểm chất lượng rượu. Vẽ biểu đồ tương quan. Nếu biết 1 giá trị alcohol hãy cho biết điểm của chất lượng rượu.

d) Sử dụng tất cả 11 thuộc tính để tìm tương quan giữa các thuộc tính với điểm chất lượng rượu. Nếu biết các giá trị tham số cho biết chất lượng rượu.