

BÀI TẬP TUẦN 5

I. Bài tập tại lớp

Bài 1: Sử dụng Python Pandas để đọc dữ liệu điểm kì thi Quốc Gia 2022 từ [Kaggle](https://kaggle.com/datasets/kuwaitcity/qg2022). Lọc ra các thí sinh dự thi ở Hà Nội, Đà Nẵng và thành phố Hồ Chí Minh.

<https://cloud.vinhthanh.net/s/P5PFPDMN8c6mFfW/download/diemthi2022.csv>

Bài 2: Tính năm số tóm lược của biểu đồ hộp ba môn toán, văn và ngoại ngữ bằng dữ liệu đã lọc từ bài 1. Sau đó vẽ biểu đồ tương ứng với năm hệ số trên trong một bản vẽ (plot) duy nhất.

Bài 3: Viết hàm nhận hai vector, tính và trả về hiệp phương sai, hệ số tương quan Pearson, hệ số tương đồng cosine, hệ số tương đồng euclidean. Vẽ bản đồ nhiệt thể hiện hệ số tương quan Pearson của ba môn toán, văn và ngoại ngữ từ dữ liệu đã lọc ở bài 1.

II. Bài tập về nhà (nhóm)

Bài 4: Đọc dữ liệu [Daily Temperature of Major Cities](https://kaggle.com/datasets/kuwaitcity/qg2022) từ link dưới đây bằng Python Pandas:

https://cloud.vinhthanh.net/s/sbFSsPr9EaCBoej/download/city_temperature.csv

Vẽ bản đồ nhiệt nước Mỹ thể hiện nhiệt độ trung bình theo từng bang từ năm 2016 đến năm 2020.

Bài 5: Kết hợp với bản đồ bài 4, vẽ thêm hai bản đồ nhiệt từ năm 2011 – 2015 và 2006 – 2010 và vẽ cả ba bản đồ trong một bản vẽ (plot) duy nhất như hình dưới đây:

