

# BÀI TẬP PANDAS

## **Bài 1:** Nhập dữ liệu từ file *hocsinh.csv*

1. In dữ liệu ra màn hình
2. In 5 dòng đầu tiên và 5 dòng cuối cùng của dữ liệu ra màn hình
3. Thống kê xem lớp có bao nhiêu bạn điểm loại giỏi (điểm từ 8 trở lên)
4. Thống kê xem lớp có bao nhiêu bạn trượt môn (điểm dưới 4 hoặc không có điểm)
5. Vẽ đồ thị histogram minh họa phân bố điểm số của lớp (trục giá trị từ 0 đến 10, không có điểm tính là 0)
6. Xử lý những ô điểm trống bằng cách fill giá trị 0 vào ô
7. Thống kê theo loại điểm, so sánh xem khác với biểu đồ histogram ở điểm nào

## **Bài 2:** Nhập dữ liệu từ file *ketqua\_soxo.csv*

1. Đọc dữ liệu từ file csv, chuyển dữ liệu cột 1 sang date
2. Xóa bỏ các dòng không có dữ liệu
3. Thêm cột mới là 2 số cuối của giải độc đắc
4. Trích xuất cột mới thành dữ liệu series để dễ xử lý
5. Hiện thị phân bố dữ liệu: biểu đồ histogram, 100 nhóm
6. Viết hàm tính số tiền thu về

## **Bài 3:** Dựa trên bộ dữ liệu xổ số, hãy thử một vài chiến lược khác dưới đây:

1. Chơi ngẫu nhiên chẵn lẻ: mỗi lần đánh cả 50 số chẵn (hoặc 50 số lẻ), chọn ngẫu nhiên
2. Chơi nuôi đầu-cuối: chọn 1 chữ số, chẳng hạn số 7, đánh cả loạt các số có đầu và cuối chứa số 7 (07,17,27,..., 97, 70,71,...,79)
3. Chơi số xuất hiện ít nhất: thống kê xem số nào xuất hiện ít nhất từ ngày đầu tiên đến trước ngày mở thưởng thì chơi số đó

## **Bài 4:** Viết chương trình Pandas để nhập dữ liệu excel đã cho (coal.xlsx) vào khung dữ liệu và tìm thông tin chi tiết trong đó 'Giờ lao động' > 20000.

**Bài 5:** Viết chương trình Pandas để tìm giá trị tổng, trung bình, tối đa, tối thiểu (sum, mean, max, min ) của cột 'Sản lượng' (Production (short tons)) của tệp coal.xlsx

**Bài 6.** Viết chương trình Pandas để nhập dữ liệu excel (coal.xlsx) vào khung dữ liệu và tìm chi tiết nơi "Mine Name" bắt đầu bằng "P"

**Bài 7.** Viết chương trình Pandas để nhập dữ liệu excel đã cho (coal.xlsx) vào khung dữ liệu Pandas và tìm danh sách các khách hàng được chỉ định theo tên.

**Bài 8.** Viết chương trình Pandas để nhập dữ liệu excel đã cho (employee.xlsx ) vào khung dữ liệu Pandas và tìm danh sách nhân viên trong đó hire\_date > 01-01-07

**Bài 9.** Viết chương trình Pandas để nhập dữ liệu excel (employee.xlsx ) vào khung dữ liệu Pandas và tìm danh sách nhân viên có ngày thuê giữa hai tháng và năm cụ thể.

**Bài 10.** Viết chương trình Pandas để nhập dữ liệu excel đã cho (employee.xlsx ) vào khung dữ liệu Pandas và chuyển đổi dữ liệu để sử dụng ngày thuê làm chỉ mục