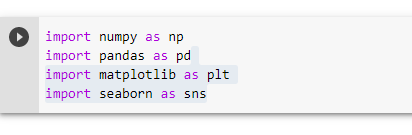
**BÀI THI**

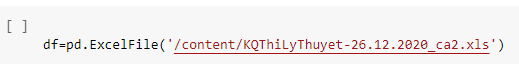
**Môn: Lập trình phân tích dữ liệu 2**

Họ và tên: Nguyễn Vũ Quốc Đại

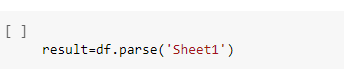
MSSV: 17063671



Khai báo các thư viện caafn thiết

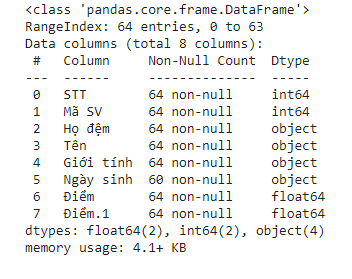


Đọc file dữ liệu excel



Lấy dữ liệu bảng Sheet1

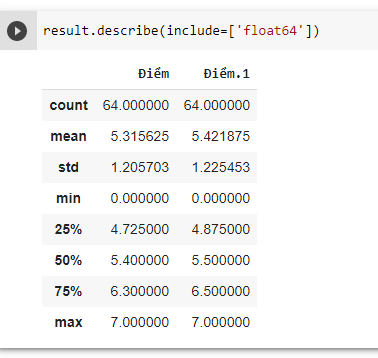




Dùng hàm info() để xem thông tin bảng dữ liệu

1.   Bảng có 8 cột

2.   Có 64 sinh viên



Dùng describe để xem các đại lượng mô tả của bảng dữ liệu:

1.   Có 64 sinh viên

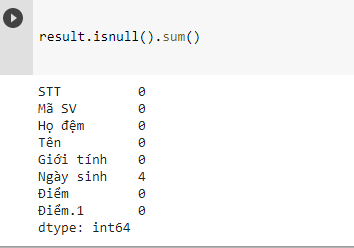
2.   Giá trị trung bình là: 5,3

3.   Độ lệch chuẩn là: 1,2

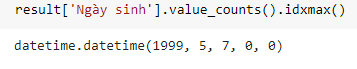
4.   Giá trị nhỏ nhất là: 0

5.   Các giá trị tứ phân vị

6.   Giá trị lớn nhất là 7



Dùng result.isnull().sum() để kiểm tra xem trong bảng dữ liệu có bao nhiêu giá trị bị thiếu. Ta thấy cột giới tính bị thiếu 4 giá trị

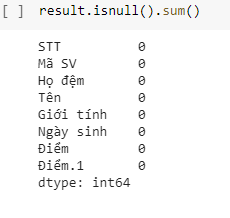


Ta tiến hành xử lý các giá trị bị thiếu:

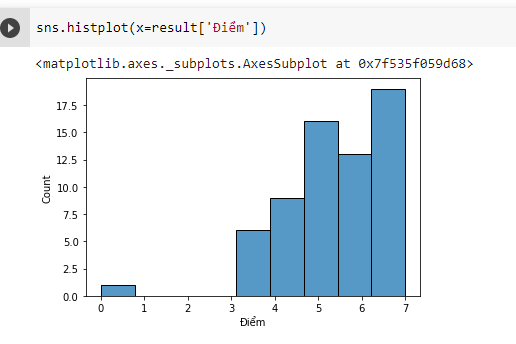
1. Bước đầu là đếm trong cột Ngày sinh có giá trị nào xuất hiện nhiều nhất



1. Ta tiến hành thay các giá trị thiếu bằng giá trị có tần xuất, xuất hiện nhiều nhất vừa tìm được



Tiến hành kiểm tra lại, ta thấy k còn cột nào bị thiếu dữ liệu

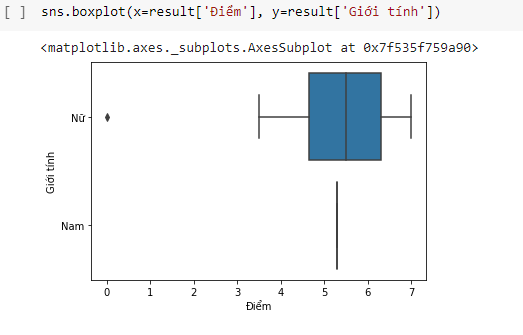


Dùng hàm histplot để thể hiện sự tương quan của biến điểm, từ biểu đồ ta thấy được rằng:

1.   Điểm 7 có số lần xuất hiện nhiều nhất, hơn 17 lần

2.   Điểm 0 xuất hiện ít nhất, khoảng 1 lần

3.   Đa số điểm tập trung nhiều ở điểm 5, 6 và 7



Dùng hàm boxplot để thể hiện sự tương quan giữa Điểm và Giới tính, ta thấy:

1.   Đa số dữ liệu đều là giới tính nữ, chỉ có 1 dữ liệu là giới tính nam

2.   Điểm thấp nhất là 0, cao nhất là 7 điểm đều thuộc giới tính nữ

3.   Đa số điểm tập trung ở điểm 4, 5, 6

4.   Có giá trị trung bình khoảng 5,3