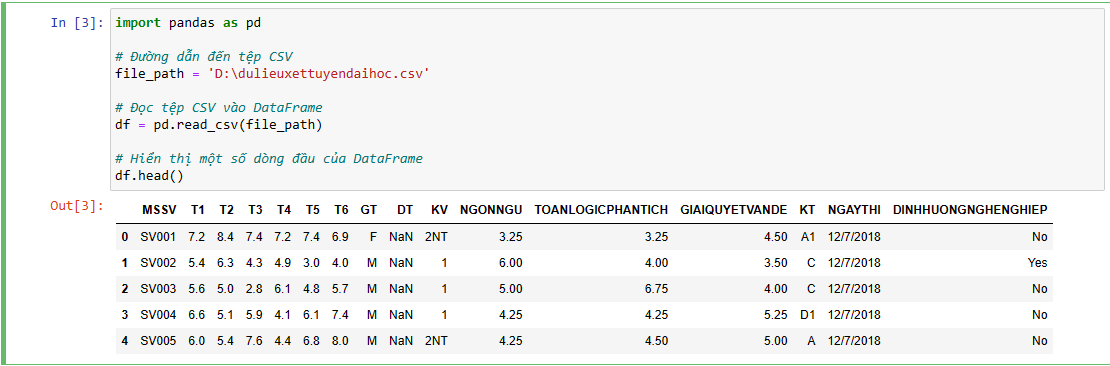
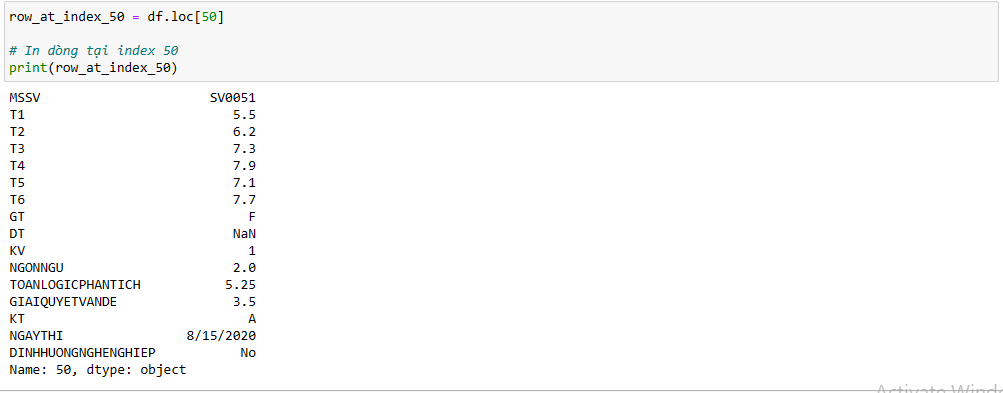
Các lệnh Python phân tích dữ liệu: Lệnh truy vấn basic

1/Import file csv cần phân tích dữ liệu

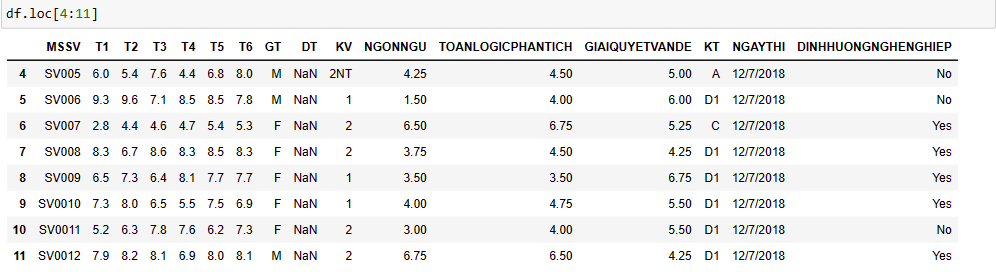


2/Hiển tự các n dòng đầu và n dòng sau: df.head(n), df.tail(n)

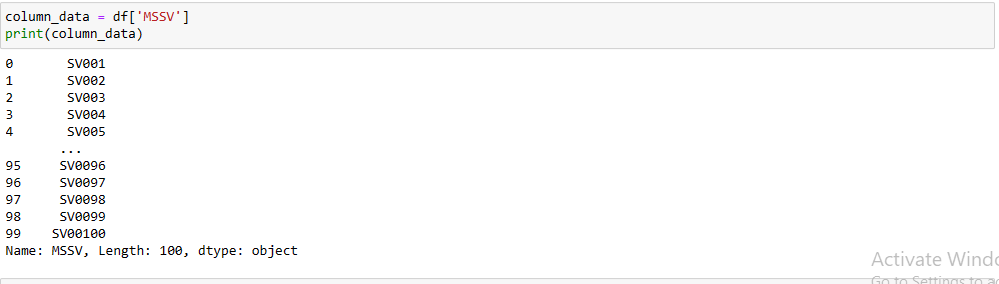
3/Xuất hàng 50



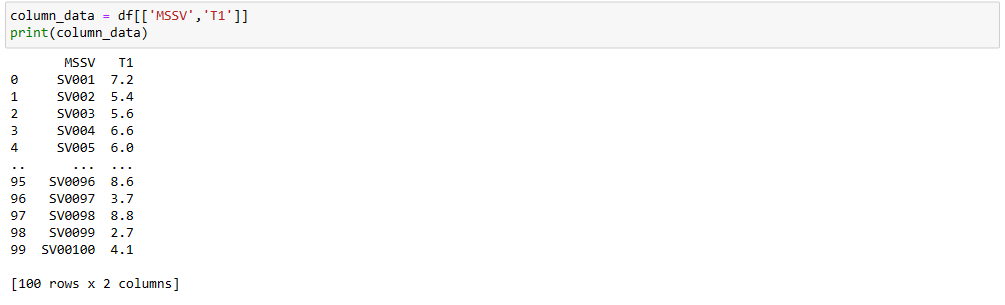
4/Xuất hàng có chỉ số 4-11



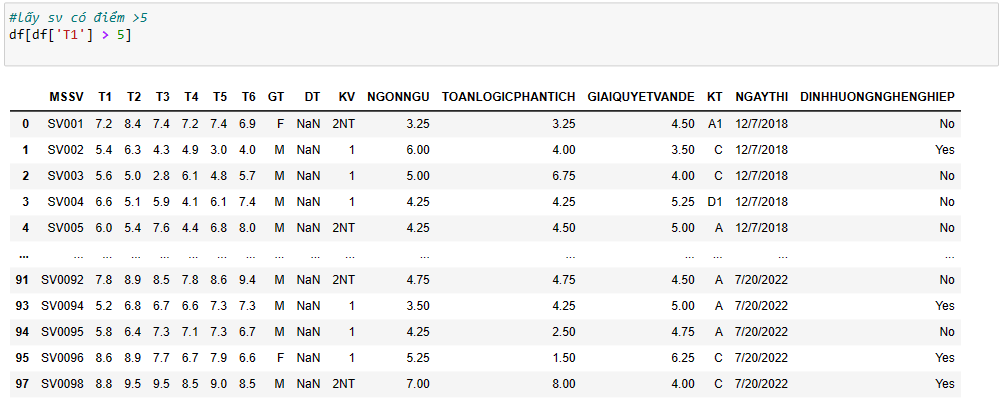
5/Xuất cột MSSV



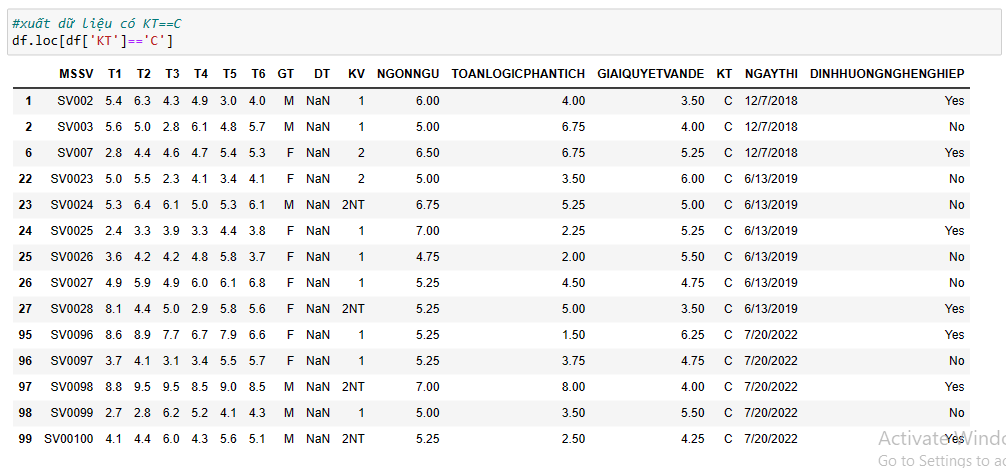
6/Xuất cột MSSV, T1 (cột khác tương tự)



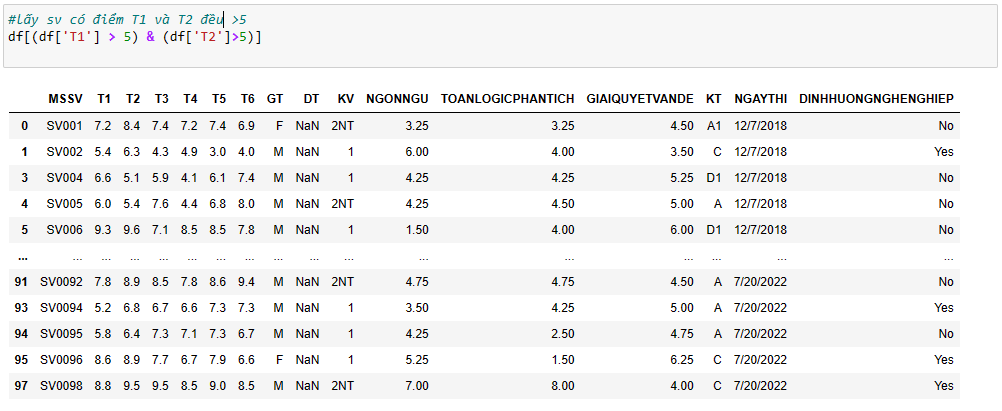
7/Lấy sinh viên có T1 >5



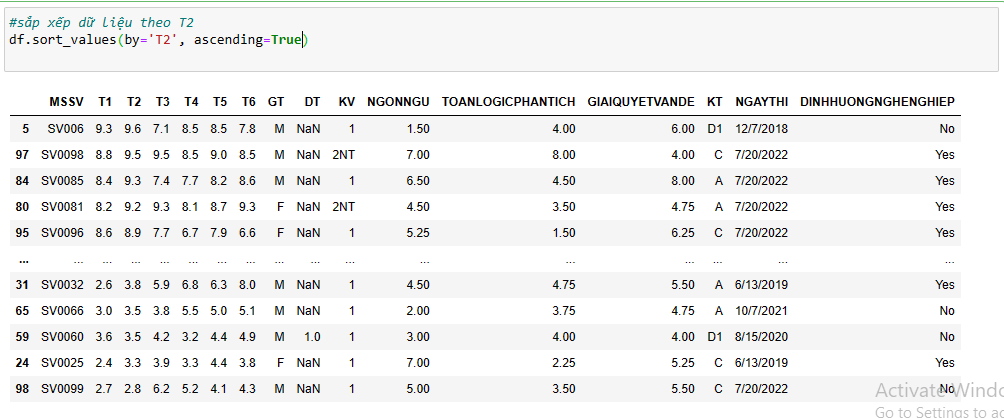
8/Xuất dữ liệu có KT==’C’



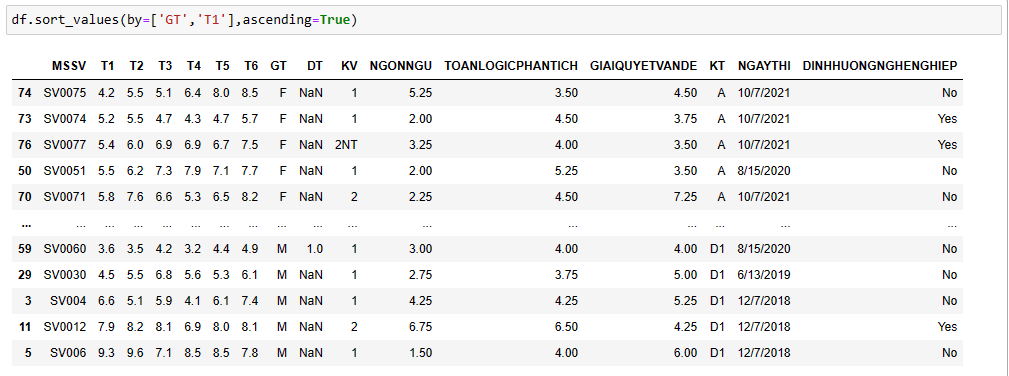
8/Lấy sinh viên có T1 và T2 lớn hơn 5



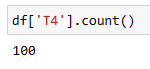
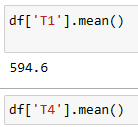
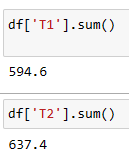
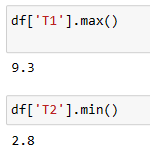
9/Sắp xếp dữ liệu theo T2 (ascending=False thì giảm dần) sắp xếp theo định lượng



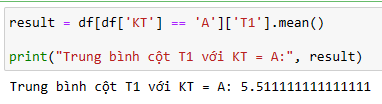
Nhóm theo định tính GT và xuất định lượng T1



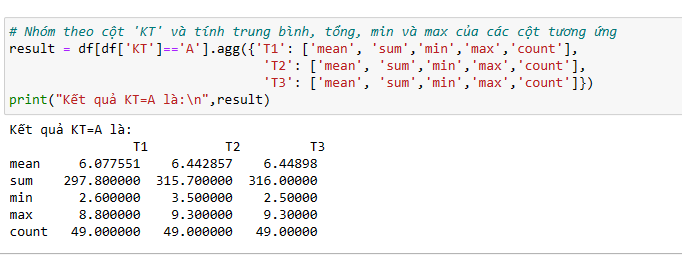
10/Min max mean sum count của **định lượng**



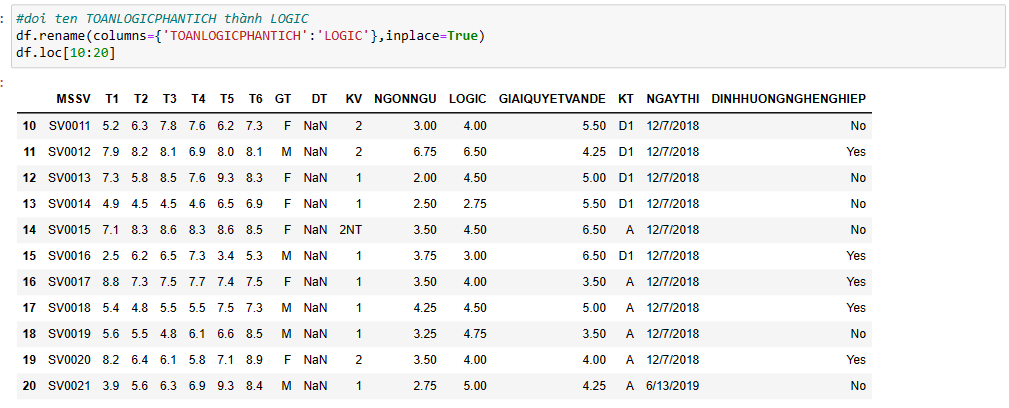
11/Trung bình T1 với KT==’A’ (xuất 1 cột)



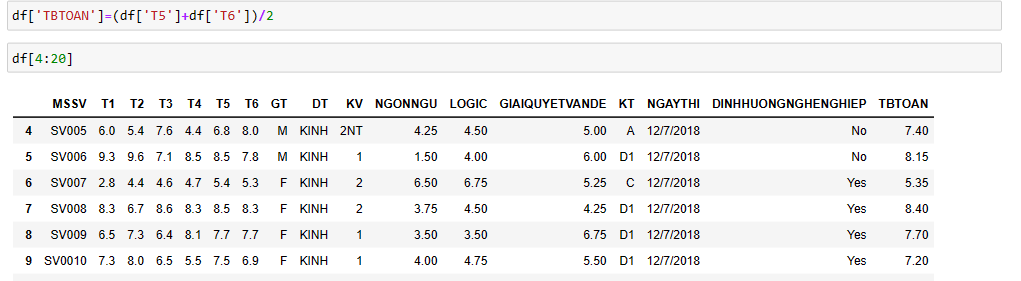
12/Nhóm KT==’A’ xuất định lượng T1 T2 T3 T4 min max mean sum count



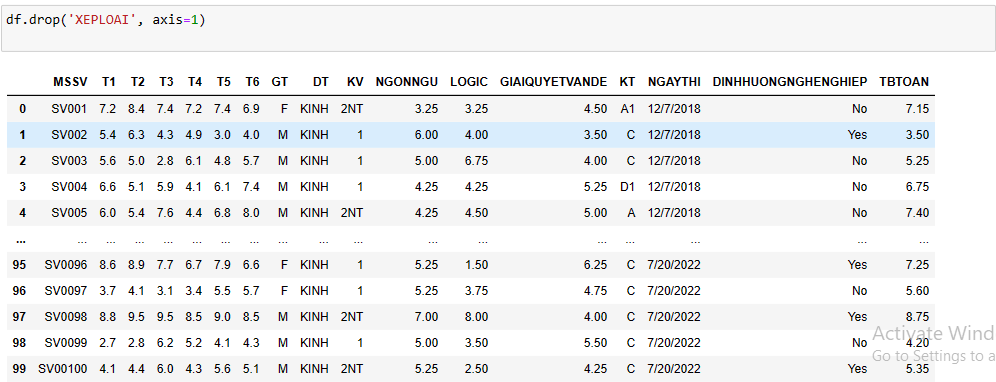
13/Đổi tên TOANLOGICPHANTICH thành LOGIC



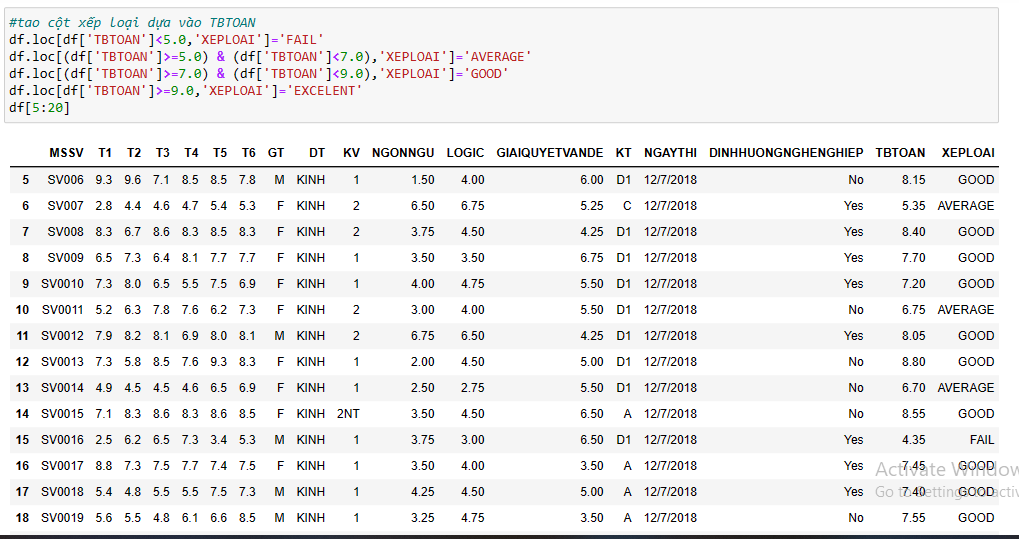
14/Tạo cột định lượng mới là trung bình lớp 12



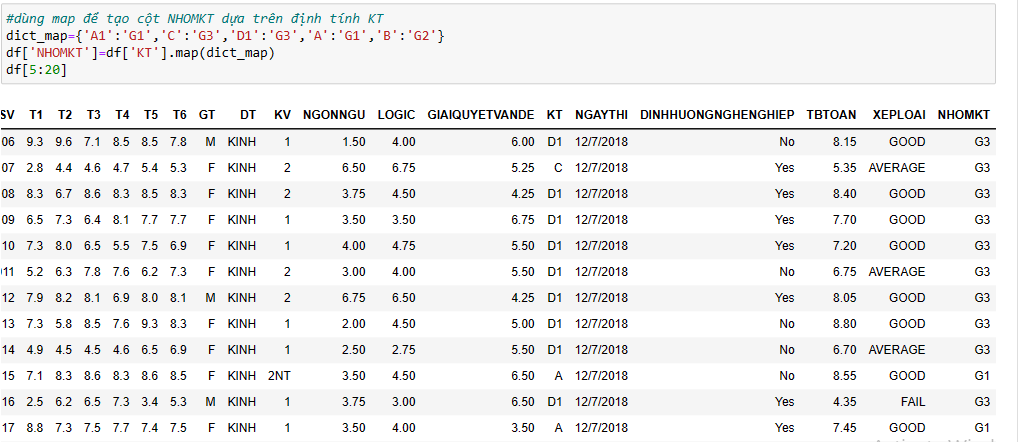
15/Xoá cột XEPLOAI



16/Thêm cột XEPHANG thoả mãn TBTOAN

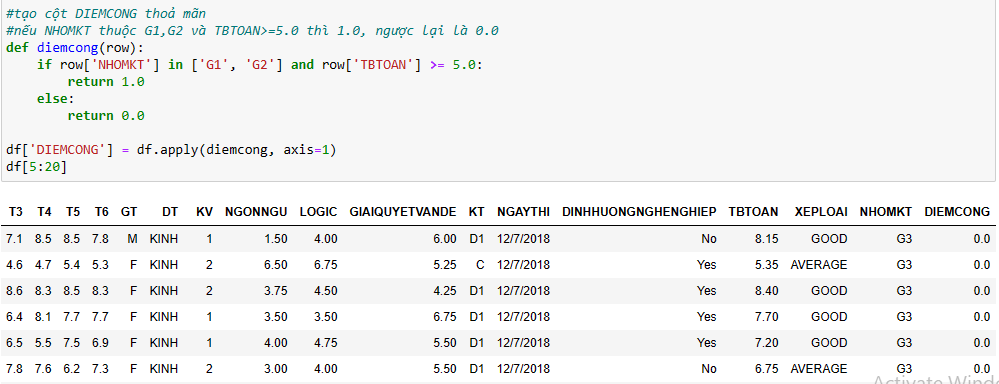


17/Thêm cột nhóm khối thi NHOMKT dựa trên định tính KT

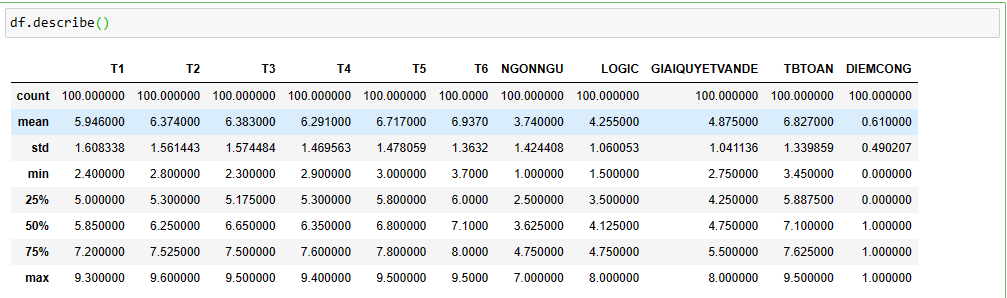


18/Tạo DIEMCONG thoả NHOMKT thuộc G1, G2 và TBTOAN>=5.0 thì 1.0, ngược lại 0.0

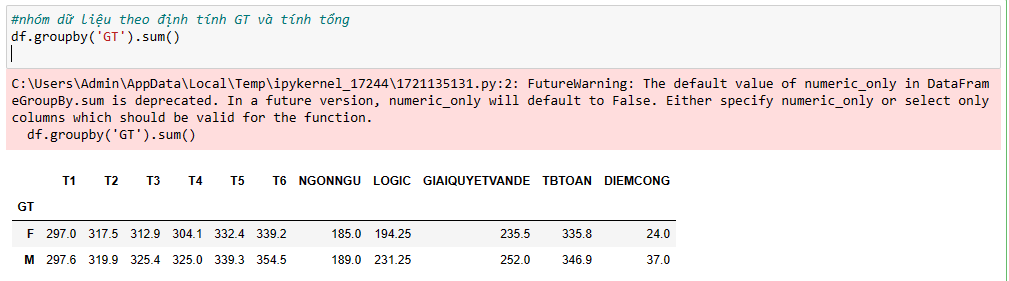
Dùng hàm def truyền tham số rồi dùng if



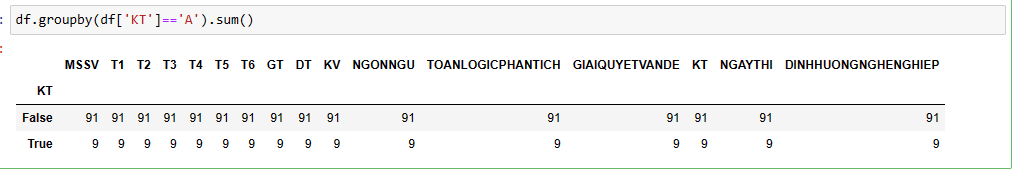
19/Tóm tắt file



20/Nhóm theo định tính và tính toàn bộ định lượng

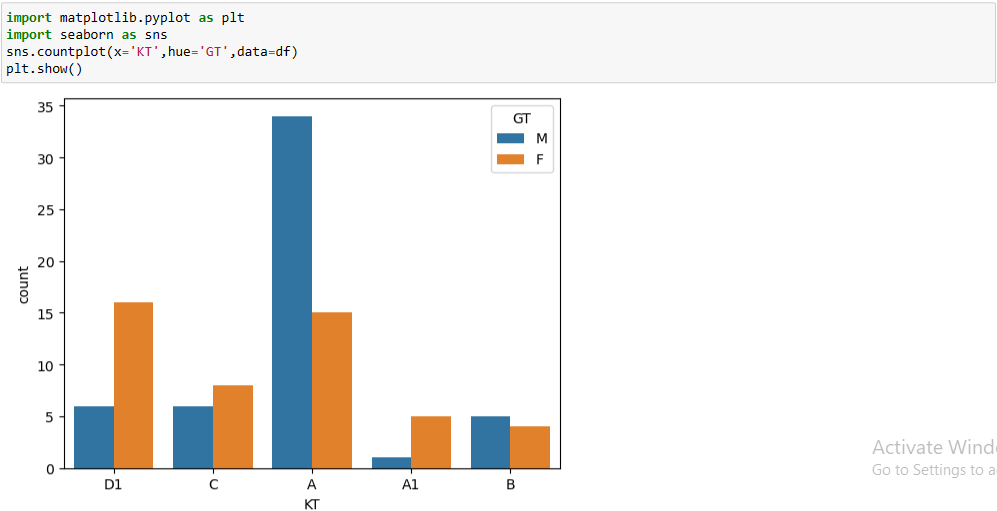


21/Nhóm theo định tính KT có ràng buộc

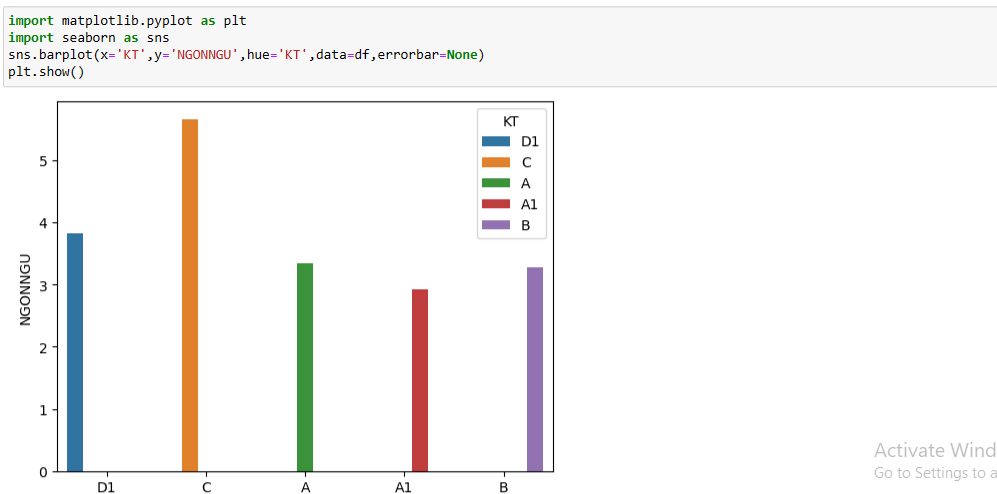


Lệnh matpotlib

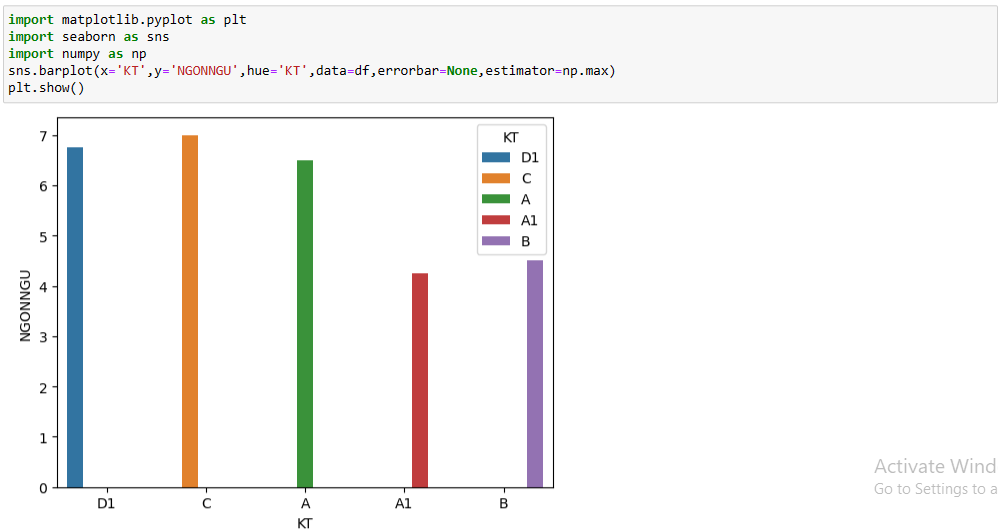
21/Vẽ biểu đồ phân tích số Giới tính (GT) theo Khối thi (KT)



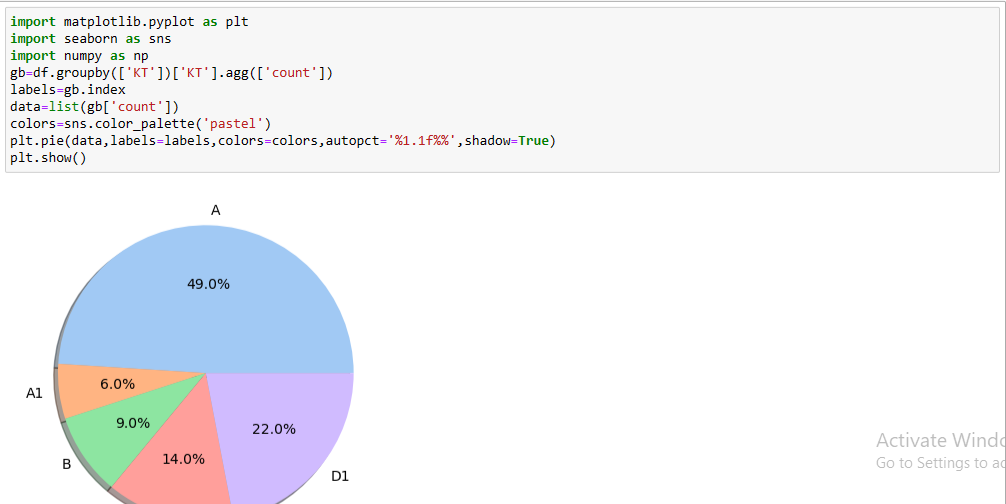
22/Vẽ biểu đồ phân tích điểm y Ngôn ngữ (NGONNGU) của Khối thi (KT) theo Giới tính (GT). Lưu ý hue=’cột định tính’



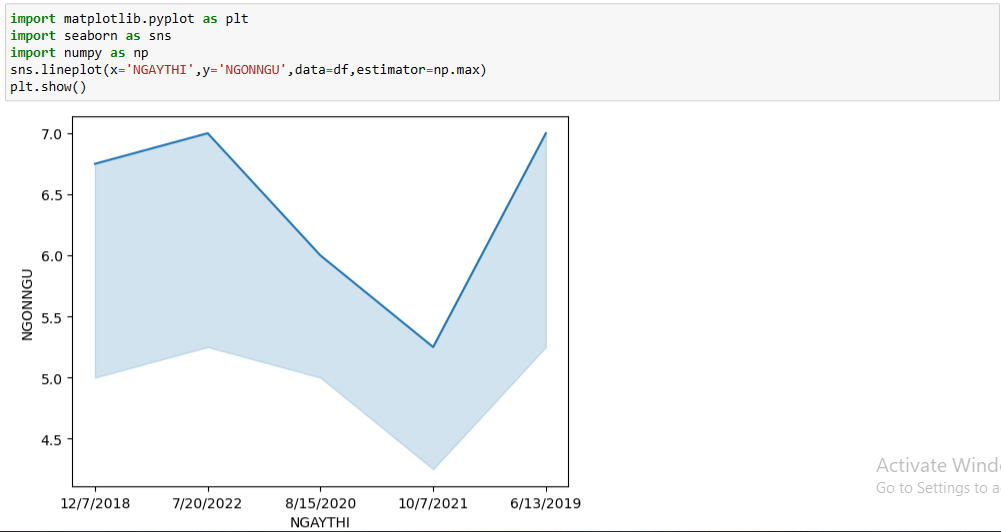
23/Nếu lấy giá trị max để so sánh



24/Tạo biểu đồ pie phân tích % khối thi (KT)



25/Dùng biểu đồ line tổng hợp trung bình điểm NGONNGU theo NGAYTHI



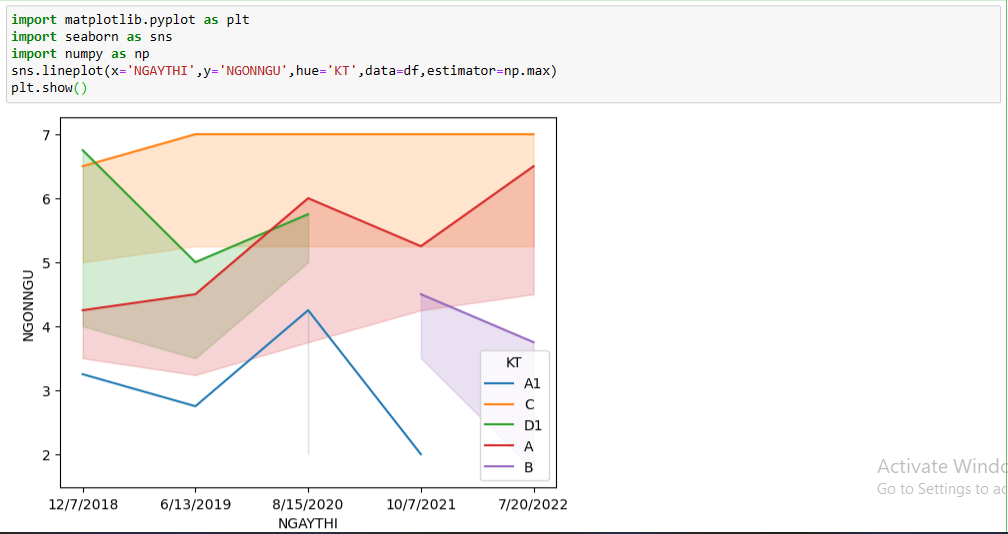
Dùng biểu đồ line tổng hợp trung bình điểm NGONNGU theo NGAYTHI

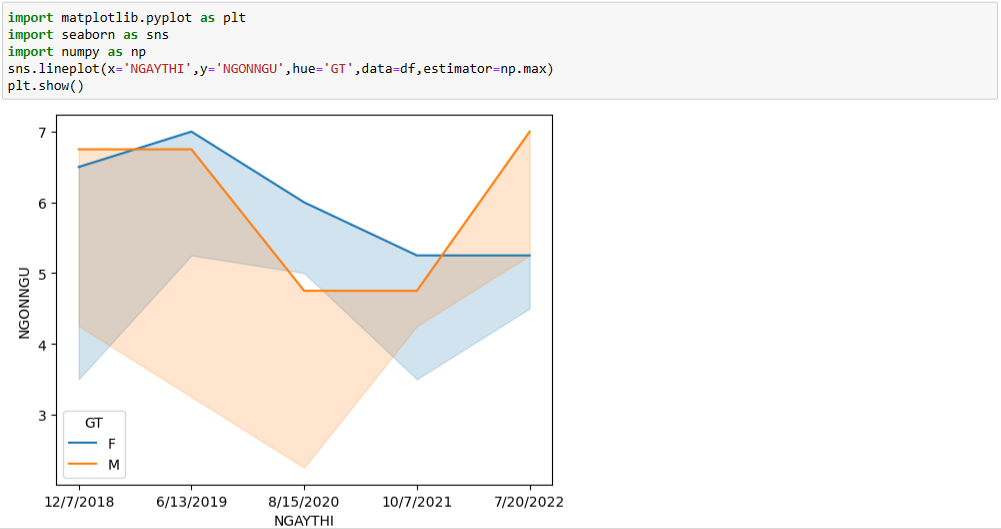
Trục x là: NGAYTHI

Định lượng cần đo là điểm NGONNGU

Đối tượng là định tính KT KhoiThi

Đối tượng sẽ được chú thích ở bảng bên





26/Điền cột DT bị thiếu



27/Tạo data frame theo ý mình

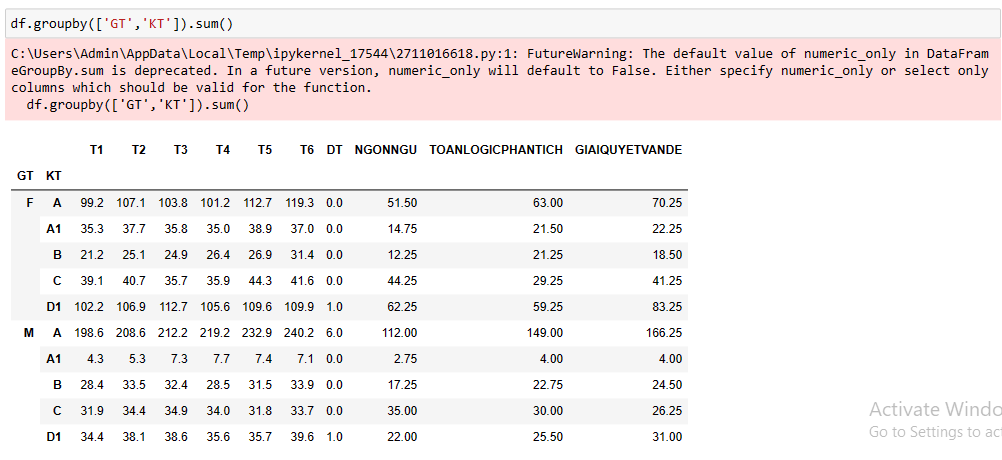
-Cột là : Name, GT, Score

-Dữ liệu bên trong [ ]

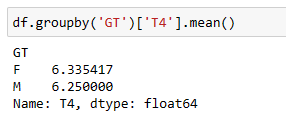
-Phải xác định định tính định lượng



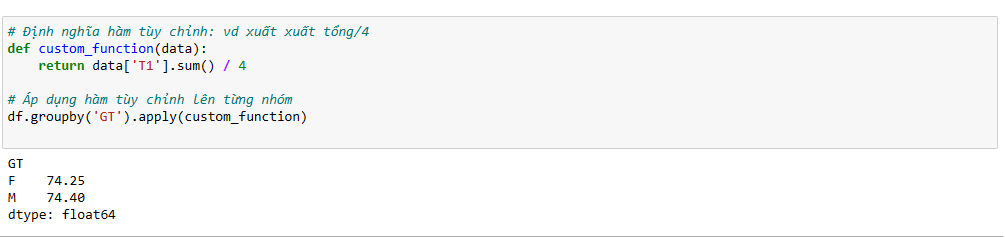
28/Nhóm nhiều cột và xuất dữ liệu định lượng



29/Nhóm GT và xuất chi tiết trung bình T4

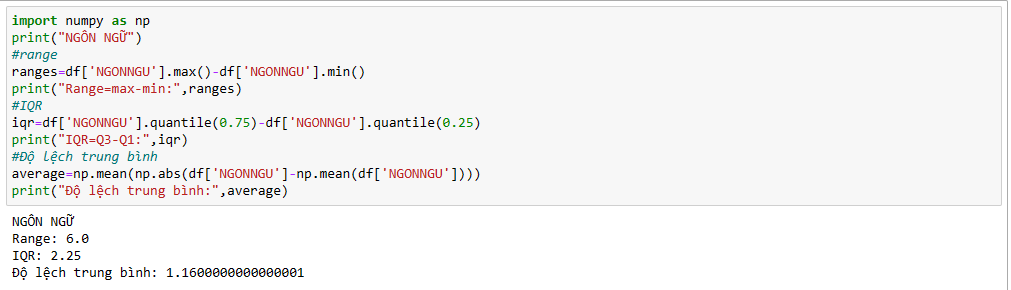


30/Nhóm GT xuất tổng T1/4



Chủ đề mô tả dữ liệu:

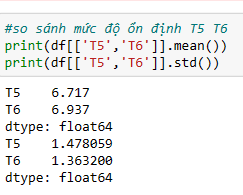
31/ Range, IQR, Độ lệch trung bình



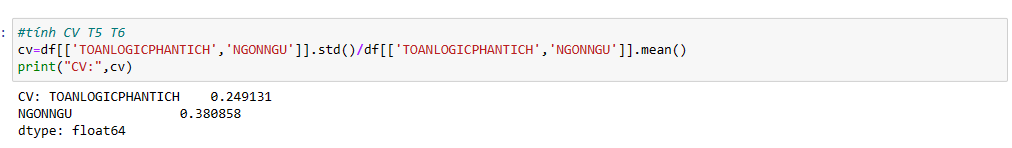
32/ Mô tả điểm NGONNGU



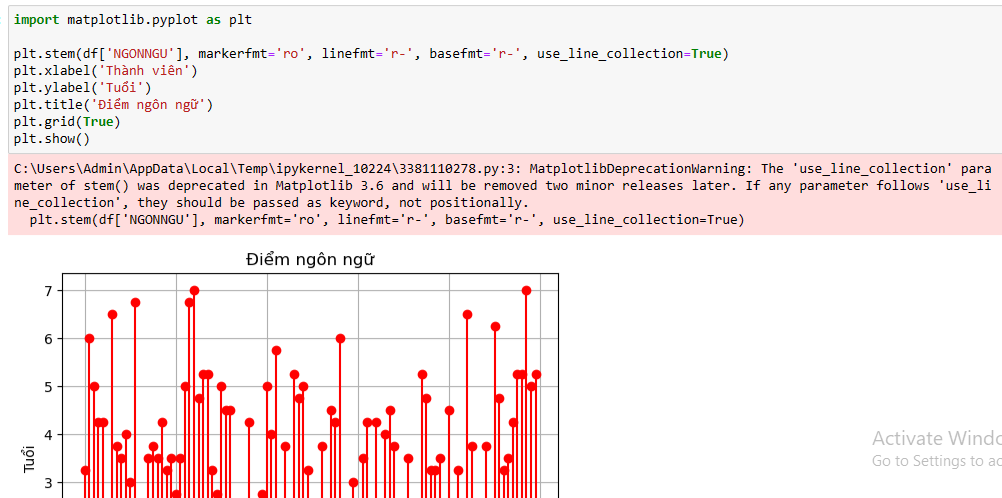
33/So sách mức độ T5 T6

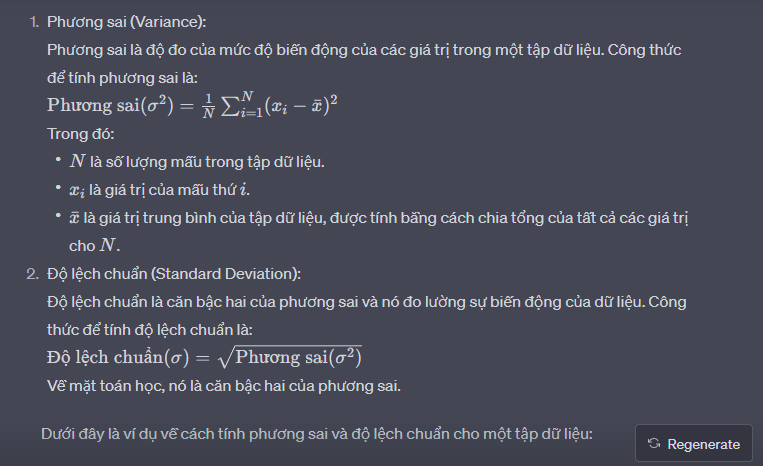


34/Tính CV cho T5 T6



35/Đồ thị thân á mô tả định lượng NGONNGU





Ví dụ khác:

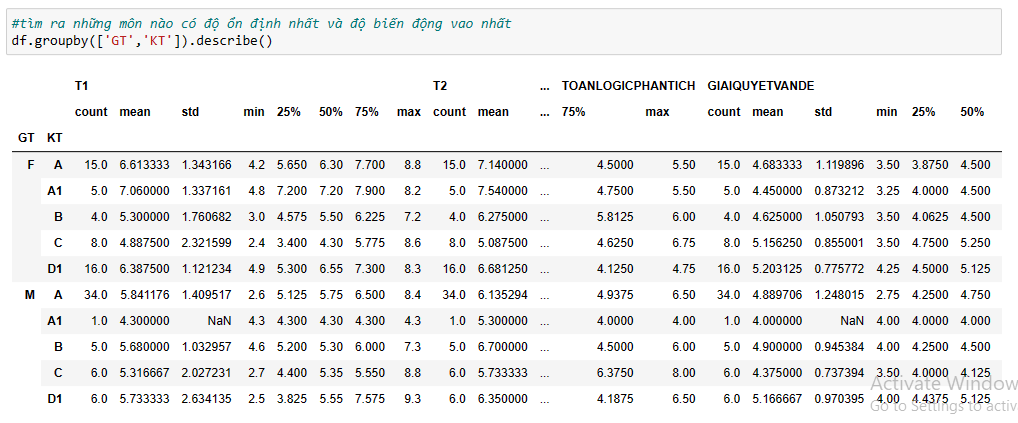
-Nhóm định tính và khối thi và mô tả các dữ liệu

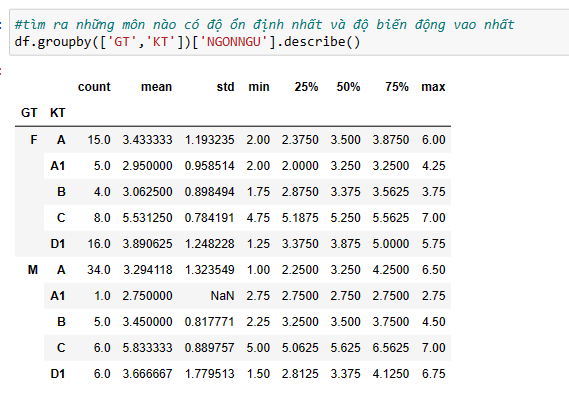
-Các mô tả của định lượng:

+Cout, sum, min, max, mean, std, var

+Phân vị Q1 Q2 Q3 với 25% 50% 75%

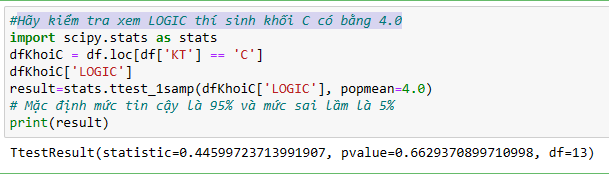
+Đề yêu cầu: Nhóm GT và KT mô tả điểm NGONNGU





PHÂN TÍCH TIN CẬY VÀ SAI LẦM CỦA ĐỊNH LƯỢNG

36/ Hãy kiểm tra xem LOGIC (định lượng) thí sinh khối C có bằng 4.0



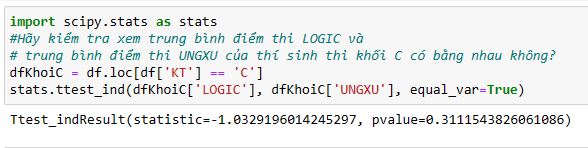
Kết luận:

Nếu pvalue > 0.05 => chấp nhận

=> LOGIC thí sinh khối C=4.0 với độ tin cậy là 95% sai lầm 5%

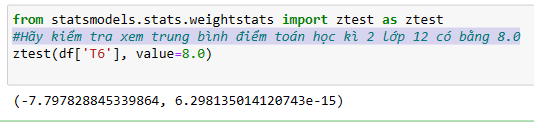
37/ Hãy kiểm tra xem trung bình điểm thi LOGIC và trung bình điểm thi UNGXU của thí sinh thi khối C có bằng nhau không?

-Lưu ý: không nói gì hết mặc định là trung bình.



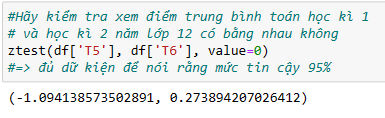
* Kết luận như trên.

38/ Hãy kiểm tra xem trung bình điểm toán học kì 2 lớp 12 có bằng 8.0 hay không (T6).

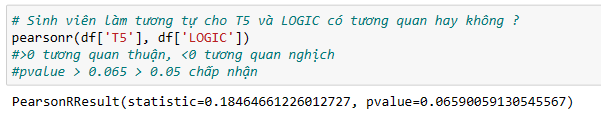


* Pvalue= 6.29x10^-15
* Không bằng 8.0

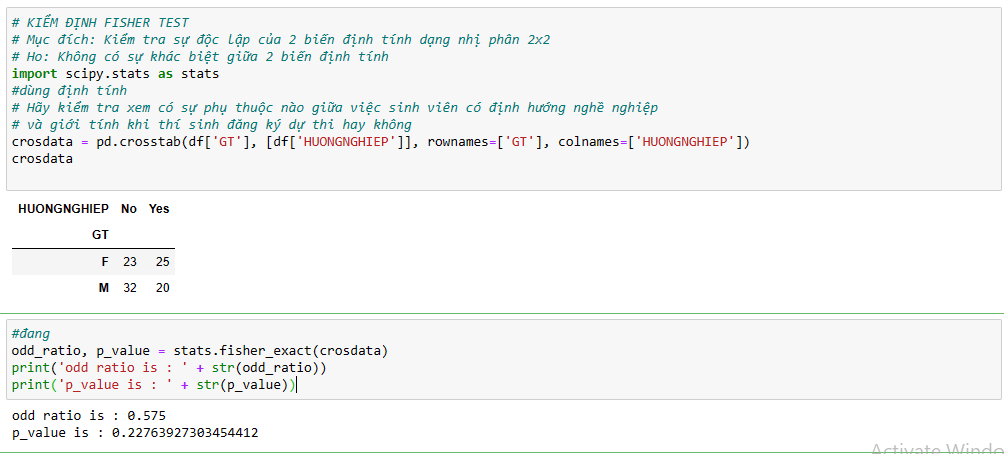
39/ Hãy kiểm tra xem điểm trung bình toán học kì 1 và học kì 2 năm lớp 12 có bằng nhau không ? => pvalue=0.27 > 0.05 => Chấp nhận.



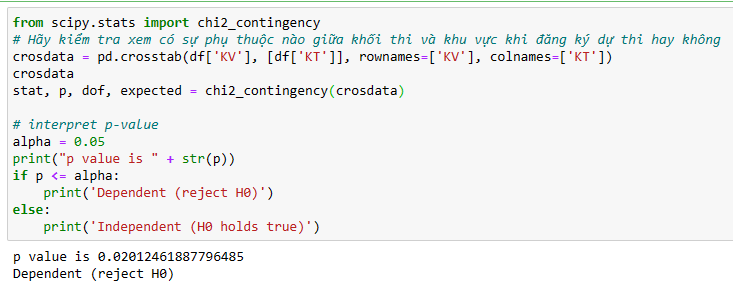
40/ Tương quan định lượng giữ 2 biến định lượng



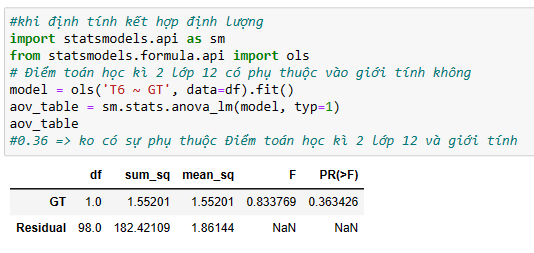
41/ FISHER TEST



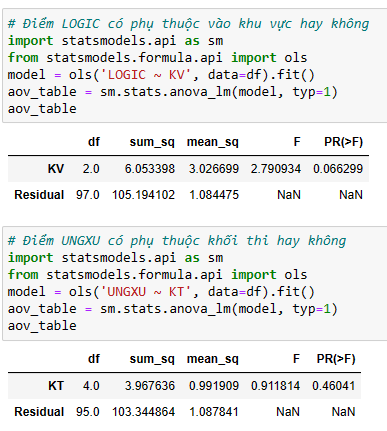
42/ SỰ PHỤ THUỘC 2 biến định lượng



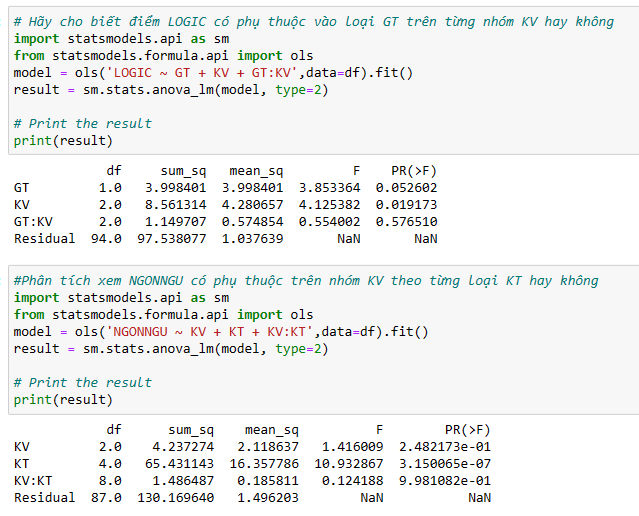
43/ Sự phụ thuộc kết hợp định tính và định lượng, xét PR của GT(định tính)



44/ 2 ví dụ tương tự, xét PR > 0.05



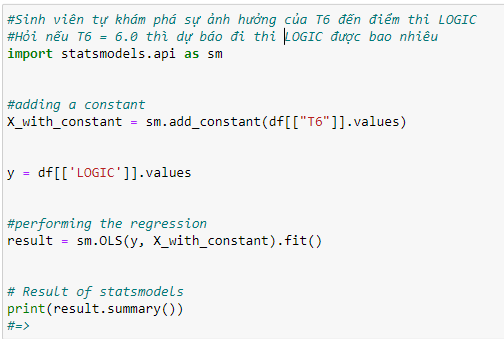
45/ Sự phụ thuộc kết hợp 2 định tính và định lượng.



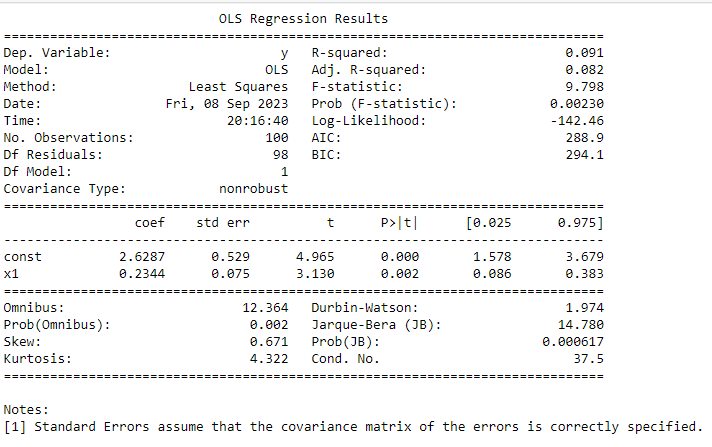
PHÂN TÍCH HỒI QUY

46/ Sinh viên tự khám phá sự ảnh hưởng của T6 đến điểm thi LOGIC

Hỏi nếu T6 = 6.0 thì dự báo đi thi LOGIC được bao nhiêu



Kết quả là:

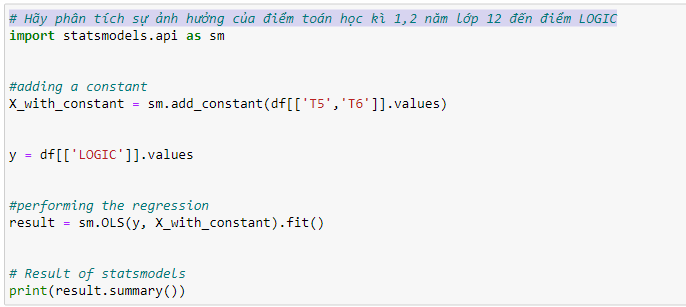


Dựa vào coef ta có phương trình hồi quy

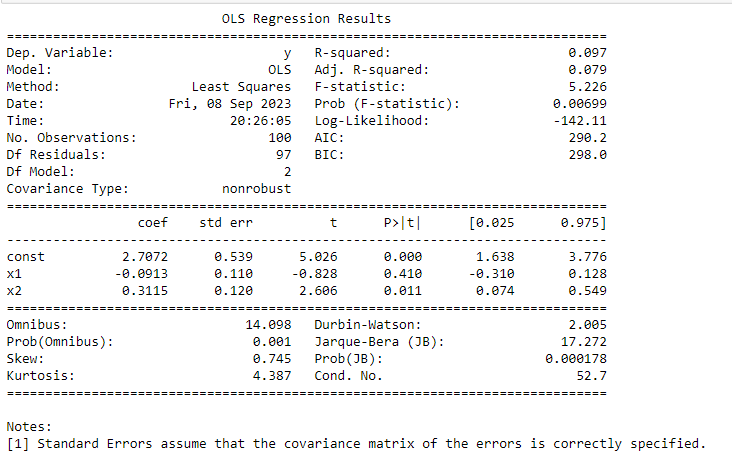
LOGIC=2.62+0.23.T6+Epsilon, đề cho T6 bằng bao nhiên thì gán T6 ra giá trị dự báo

Từ giá trị thực tế ta suy ra Epsilon

47 Hãy phân tích sự ảnh hưởng của điểm toán học kì 1,2 năm lớp 12 đến điểm LOGIC



Kết quả:

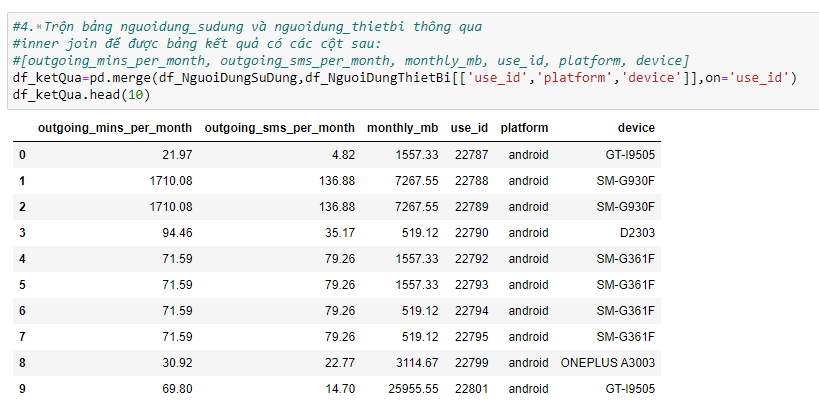


Dựa vào coef ta có phương trình hồi quy

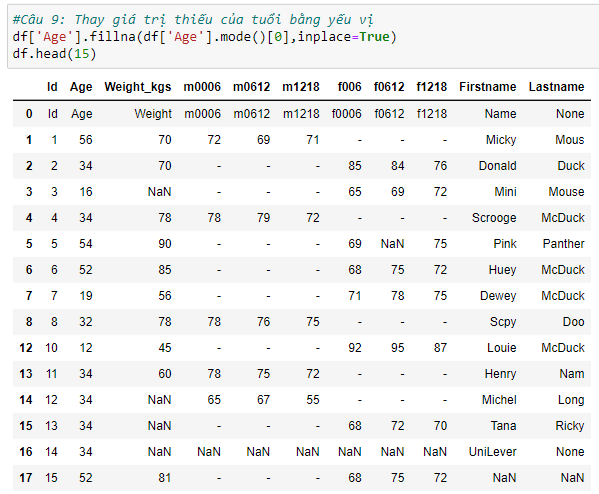
LOGIC=2.62+0.23.T6+Epsilon, đề cho T6 bằng bao nhiên thì gán T6 ra giá trị dự báo

Từ giá trị thực tế ta suy ra Epsilon

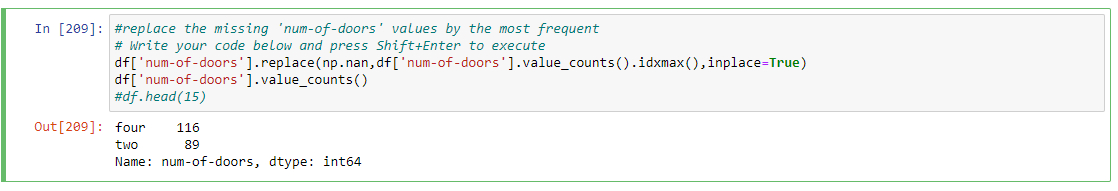
48/ Sát nhập các cột của bảng dữ liệu



49/ Điền cột định lượng rỗng

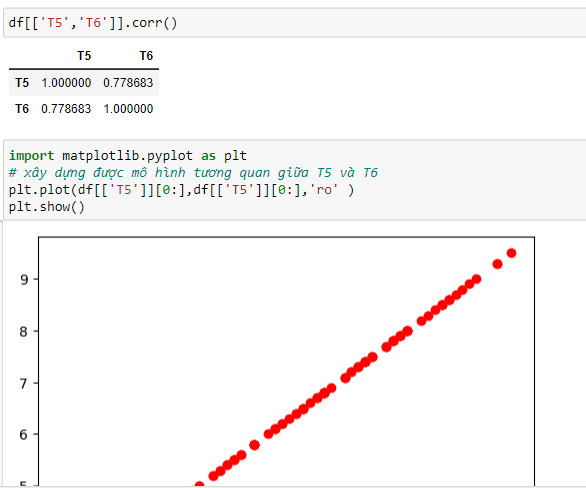


50/ Điền cột định lượng bằng giá trị định lượng xuất hiện nhiều xuất rồi thống kê cout định lượng.

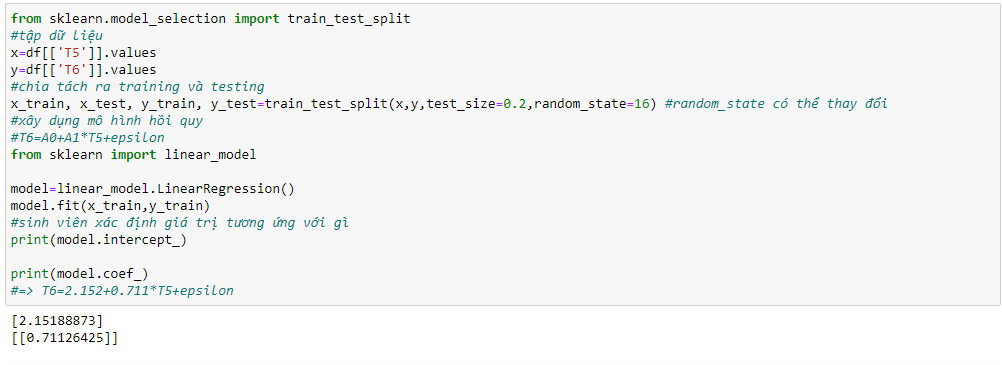


MACHINE- LEARNING

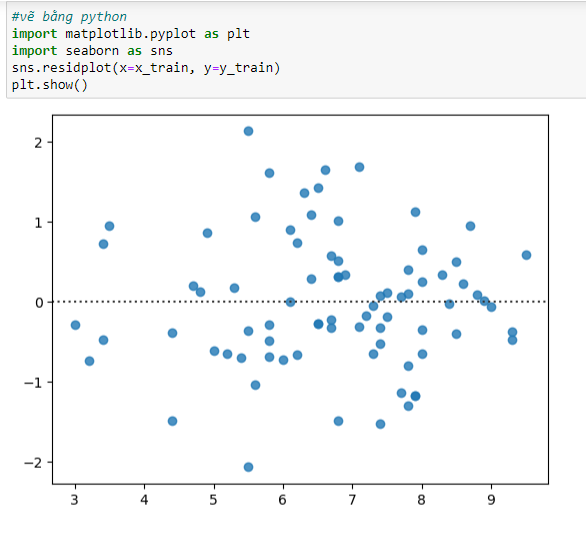
51/ Sự tương quan điểm T5 và T6



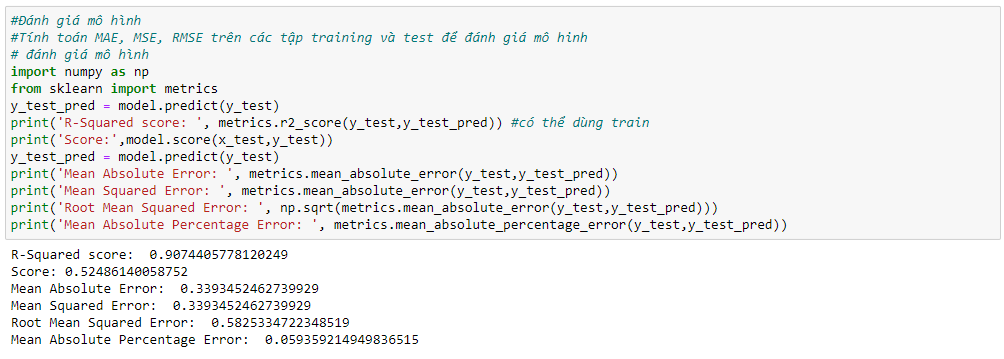
52/ Xác định A0 và A1 cho phương trình hồi quy



53/ Vẽ biểu đồ scatter biểu diễn giá trị x\_train, y\_train



54/ Tính MAE, MSA, RMSA trên tập traning và test



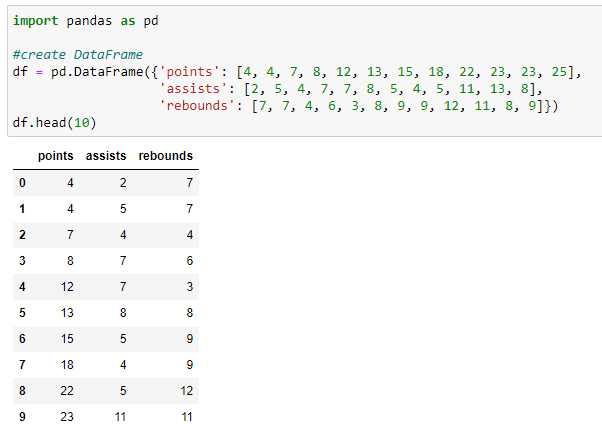
Lưu ý: có thể thay thế tập traning thay test

55/ Lưu trữ mô hình thành file sav và đọc file đó để dự báo.

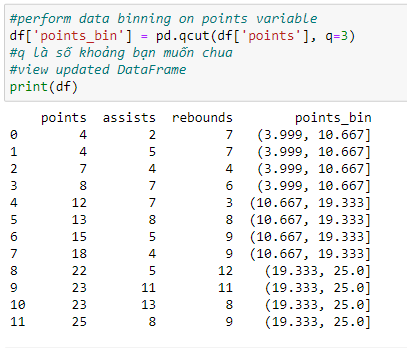


BINNING DATA (giữa 2 định lượng)

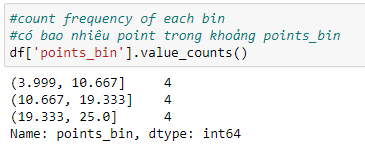
Nhắc lại tạo data frame theo ý mình



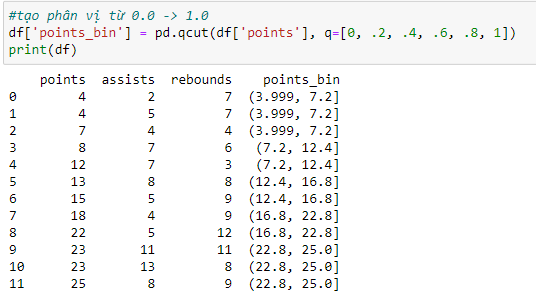
56/ Tạo point-bin lấy dữ liệu point để phân vị các khoảng muốn chia q=3 là 3 khoảng

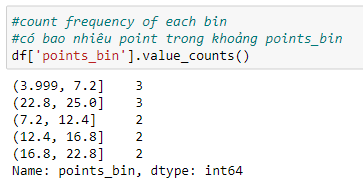


57/ Thống kê point theo point-bin

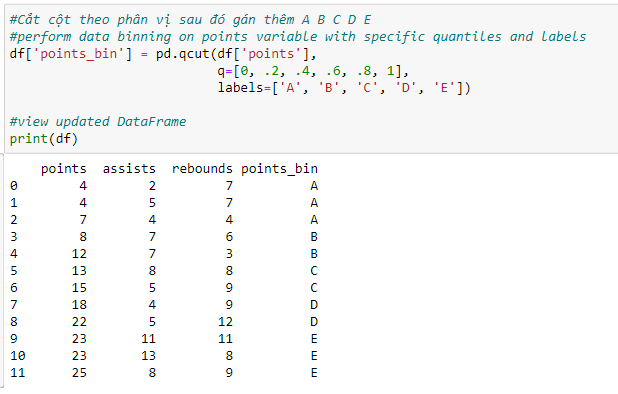


58/ Tạo phân vị theo ý mình. Ví dụ 0-1 thì step là 0.2 và thống kê.

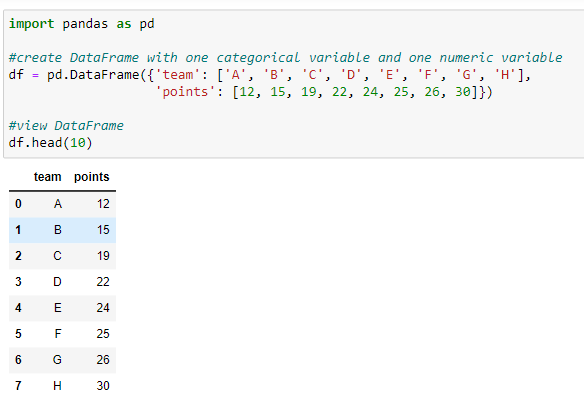




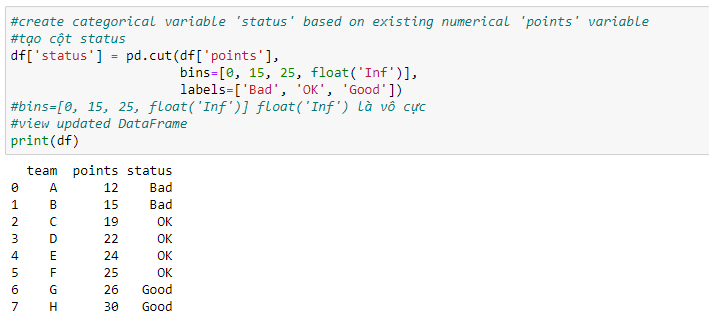
59/ Cắt phân vị thay thế nó thành A B C D E



60/



Yêu cầu: tạo cột status xét point, phân vị theo điều kiện (float(‘Inf’) là dương vô cực)



61/ Tạo data frame theo mảng và gán tiêu đề cột.



Kĩ thuật số hoá thành 1 0. Giải thích:

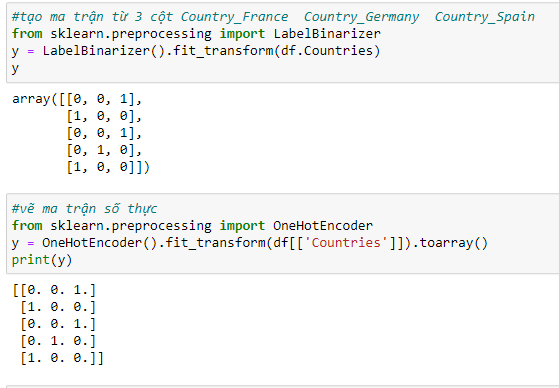
-Định lượng được chia thành nhóm nhỏ

-Mỗi hàng thì thuộc định lượng nào

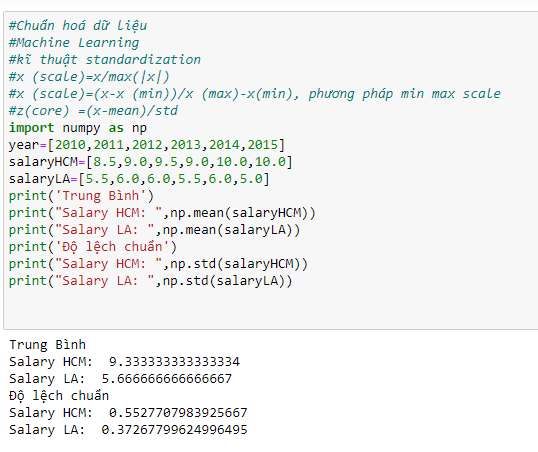
-Đánh số 1 vào cột có nhóm đó



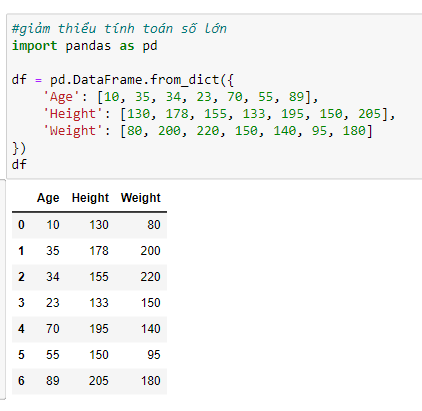
62/ Ma trận số hoá

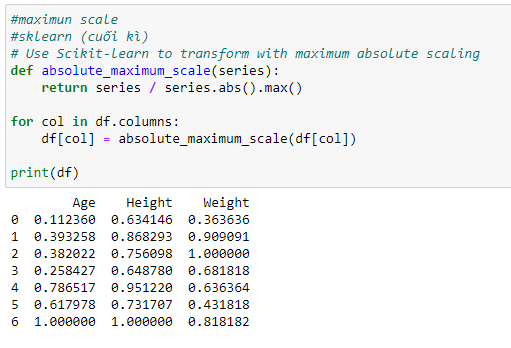


63/ Tính trung bình và độ lệch chuẩn 2 định lượng lương HCM và lương LA

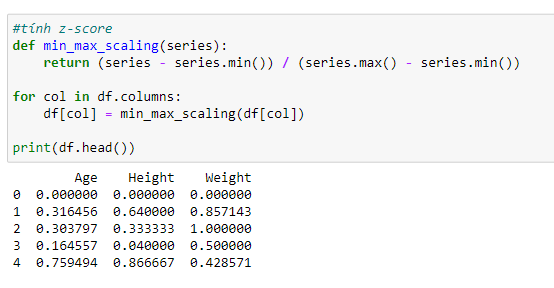


64/ Tạo data frame và tính x (scale)=x/max(|x|)

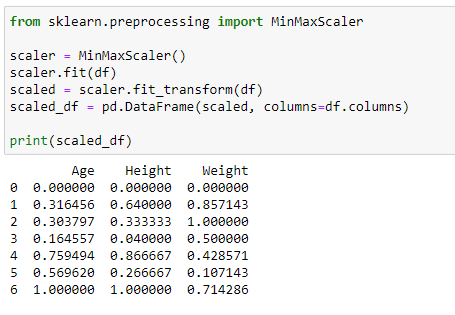




Tính z-score



Tính z-score cách 2



65/ Vẽ biểu đồ làm mịn dữ liệu

